

# Psychological Bulletin

HARRY HELSON, Editor

*University of Texas*

---

## CONTENTS

The Second Facet of Forgetting: A Review of Warm-Up Decrement	JACK A. ADAMS	257
On the Reformulation of Inhibition in Hull's System	ARTHUR R. JENSEN	274
Acquiescence and the Factorial Interpretation of the MMPI	SAMUEL MESSICK AND DOUGLAS N. JACKSON	299
Scales and Statistics: Parametric and Nonparametric	NORMAN H. ANDERSON	305
Basic Forms of Covariation and Concomitance Designs	RICHARD W. COAN	317
The Self-Concept: Fact or Artifact?	C. MARSHALL LOWE	325

---

Published Bimonthly by the  
American Psychological Association

VOL. 58, No. 4

JULY 1961

### Consulting Editors

W. BEVAN, JR.  
*Kansas State University*

R. R. BLAKE  
*University of Texas*

W. R. GARNER  
*Johns Hopkins University*

J. P. GUILFORD  
*University of Southern California*

W. A. WILSON, JR.  
*Bryn Mawr College*

W. H. HOLTMAN  
*University of Texas*

L. J. POSTMAN  
*University of California, Berkeley*

J. B. ROTTER  
*Ohio State University*

S. B. SELLS  
*Texas Christian University*

The *Psychological Bulletin* contains evaluative reviews of research literature and reviews of research methodology and instrumentation in psychology. This JOURNAL does not publish reports of original research or original theoretical articles.

Manuscripts should be sent to the Editor, Harry Nelson, Department of Psychology, University of Texas, Austin 12, Texas.

*Preparation of articles for publication.* Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1957 Revision). Special attention should be given to the section on the preparation of the references (pp. 50-60), since this is a particular source of difficulty in long reviews of research literature. All copy must be double spaced, including the references. All manuscripts should be submitted in duplicate, one of which should be an original typed copy; author's name should appear only on title page. Dittoed and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

*Reprints.* Fifty free reprints are given to contributors of articles and notes.

ARTHUR C. HOFFMAN  
*Managing Editor*

HELEN ORE  
*Promotion Manager*

Communications—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street N.W., Washington 6, D.C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Annual subscription: \$10.00 (Foreign \$10.50); Single copies, \$2.00.

PUBLISHED BIMONTHLY BY:  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin

and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Washington, D.C., and at additional mailing offices. Printed in U.S.A.

Copyright The American Psychological Association, Inc., 1961.

# Psychological Bulletin

## THE SECOND FACET OF FORGETTING: A REVIEW OF WARM-UP DECREMENT

JACK A. ADAMS<sup>1</sup>

*University of Illinois*

The interference theory of forgetting assumes that the extraexperimental occurrences of S-R sequences, either before original learning of goal responses<sup>2</sup> or interpolated between original learning and recall, will induce a decrement at recall if their stimuli are the same or similar to those of the criterion task and responses are antagonistic. Thus, the laws of forgetting reduce to the laws of proactive and retroactive inhibition (Briggs, 1957; Bugelski & Cadwallader, 1956; Osgood, 1949), with experimental extinction as the process whereby responses are weakened in interference paradigms (Adams, 1952a; Briggs, 1954; Underwood, 1948a, 1948b; Underwood & Postman, 1960). More recently, Underwood (1957) has shown the potency of proactive inhibition on the recall of verbal responses by demonstrating that the prior learning of verbal materials has led us to greatly overestimate the amount forgotten. Underwood's 1957 study, combined

with recent research by Underwood and Postman (1960) showing effects on verbal recall from expected sources of verbal interference outside the laboratory, have materially strengthened the interference theory of forgetting. Additional evidence for the interference theory has been by Steinberg and Summerfield (1957) and Summerfield and Steinberg (1957, 1959) who used nitrous oxide in the control of learned associations during interpolated rest. Osgood (1953, pp. 593-597) presents a good review of research in support of the interference theory where various techniques are used to control activities of the organism during the retention interval as a means of reducing opportunities for learning competing responses. Other explanations of normal forgetting might eventually be shown to have validity also, but the preponderance of contemporary evidence lies in support of the interference theory and it will be used without further qualifications throughout this paper as the mechanism by which goal responses are directly influenced and weakened during a retention interval.

The purpose of this paper is to review evidence for the view that warm-up decrement (WU) is a second portion of the retention loss, arising from conditions other than direct interference with goal responses. Irion

<sup>1</sup> Several psychologists read a draft copy of this paper and improved it with their thoughtful commentary. Acknowledgement is due to A. M. Barch, R. C. Davis, M. R. Denny, J. M. Digman, C. P. Duncan, J. C. Jahnke, and B. J. Underwood.

<sup>2</sup> The term "goal response" is used throughout this paper as synonymous with "test response" or " criterial response." It is a feature of overt behavior which the experimenter records and uses as the dependent variable.

(1948) points out that the very special circumstances of stimulus and response similarity required for interference, along with the amount of interfering activity required to produce decrement in the originally learned responses, make it unlikely that the fortuitous experiences of everyday life outside the laboratory could induce significant amounts of one-factor forgetting. While the work of Underwood and Postman (1960) suggests that casual interference is a factor to be reckoned with, there is the strong prevailing sentiment in experimental psychology, supported by research evidence, that hypothesizes WU as a second part of forgetting independent of direct interference with the goal responses. As we shall see, the support for this two-factor view is not as secure as it might be.

#### HISTORICAL BACKGROUND AND DEFINITIONS

The first systematic observations on WU arose from interest in fatigue and the characteristics of performance curves under conditions of protracted work, and they appear to have been made in the latter part of the 19th century by Kraepelin and his students (Arai, 1912). Studying a variety of tasks, these researchers observed that the initial segment of a performance curve was typified by a rapid rise in efficiency, followed by a much slower rate of increase or a decline when fatigue effects were present. They identified this initial rapid rise as WU, although in some cases it could have been considered a practice effect or simple reacquisition following one-factor forgetting. Interestingly, these early investigators made an observation which enters the thinking of many later workers: that a rest period contains the simultaneous and opposing processes of bene-

ficial recovery from decremental work effects, and loss of advantageous factors whose reinstatement occurs during the warming-up period.

Mosso (1906) reported anecdotal accounts on the need for poets and writers to warm-up before a period of productive work could begin. Wells (1908) observed the rapid increase in initial postrest performance on a tapping test which, by this time, generally had become identified as WU. Thorndike (1914) in a chapter "Mental Work and Fatigue" gives a more careful definition than previous investigators:

The best definition of "warming-up" as an objective act is that part of an increase of efficiency during the first 20 minutes (or some other assigned early portion) of a work period, which is abolished by a moderate rest, say of 60 minutes (p. 66).

One other quotation from Thorndike is particularly significant:

It should also be noted that intellectual warming-up in the popular sense refers rather to fore-exercise of *other functions*, in order to get materials and motives with which and by which the given function is to work, than to an intrinsic alteration of it (pp. 67-68).

Thorndike's definition of WU as a rapid increase of efficiency during the initial postrest period is consistent with that of earlier writers. Thorndike, in these quotations, makes the influential observation that the WU segment is something other than strengthening of goal responses with practice, and he clearly makes this point in the second quotation (pp. 67-68) when he identifies intellectual WU as the fore-exercise of other functions. It is this identification of WU with factors supporting goal responses which stands as the foundation of the two-factor theory of forgetting, and apparently Thorndike was the first to make it.

Thorndike's observations stand as

the most important historical predecessors of contemporary views, but other investigators made observations on warm-up too, with an occasional experiment. Watson (1919, pp. 354-355) assumed that WU appeared only for heavy muscular work and that the warming-up period was a time of increased glandular action. Robinson and Heron (1924) defined WU as "a rise in efficiency which is steeper and more temporary than the rise which can be seen, let us say, in successive daily performances" (p. 81). Robinson (1934) essentially repeated his 1924 views. Snoddy (1935) presented the first data from a relatively large group of subjects which showed WU following rest. He employed a mirror-tracing instrument as his experimental device. Bell (1942) performed an experiment on the Rotary Pursuit Test (Melton, 1947) on the effects of varying amounts of rest interpolated early and late in practice. Warm-up decrement, as measured by the difference between the first and second postrest trial, was found to first increase and then decrease with amounts of interpolated rest ranging from 1 minute to 30 hours. This trend applied to both early and late in practice.

Post-World War II research displayed an accelerated interest in WU and produced more careful definitions, hypotheses concerning its underlying nature, and specific experimental tests. The modern investigators generally followed the leads of their predecessors. Ammons (1947a), in a miniature system of variables determining rotary pursuit performance, measured WU as the difference between the score on the first postrest trial and a point on the performance curve estimated as the level that would have occurred had there

been no need for warming-up. Irion (1948, p. 338) defines WU on the response side in terms of the greater slope of the initial segment of the postrest curve relative to the slope of the original learning curve at a corresponding level of initial proficiency. The response definitions by Ammons and by Irion amount to about the same thing and, along with their theoretical views to be discussed subsequently, have been the mainstay of most workers in the area since the war. A significant feature of these definitions is that they do not imply an actual decrement from the last prerest trial to the first postrest trial and, in this sense, the common reference to "decrement" is a misnomer. Consistent with most early observations on WU, the definitions involve an expression of the sharp initial rise in a postrest performance curve and are independent of whether there is an overall gain or a loss over rest. It is a decrement only in the sense that initial postrest performance is below an expected level because of WU, and this expected level is not always below the level on the final postrest trial. This is the interaction of work and WU effects over rest which drew the attention of Kraepelin and his associates (Arai, 1912). Figure 1 illustrates WU and its appearance under conditions of massed and distributed practice (from Adams, 1952b). The Rotary Pursuit Test was used and 5 days of practice were administered, with 36 ten-second trials given each day. Massed practice was 6 minutes of continuous practice, and distributed practice had a 40-second intertrial rest interval. Eighteen subjects were in the massed group and 21 in the distributed group. These data are a good example of WU manifestations, although a subsequent section

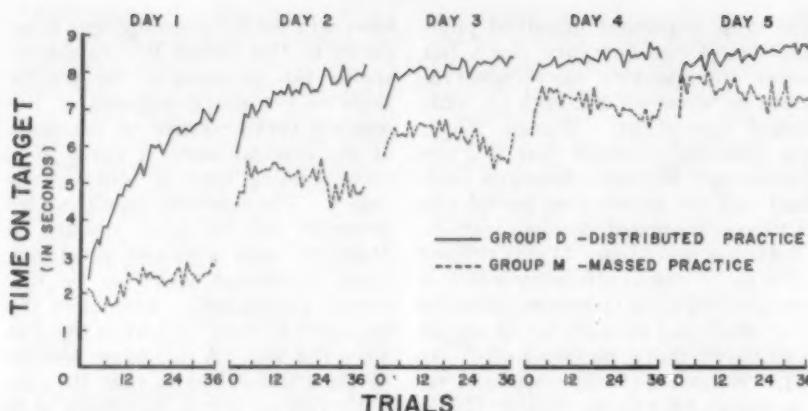


FIG. 1. Illustrations of warm-up decrement under conditions of massed and distributed practice on the Rotary Pursuit Test. (From Adams, 1952b)

will point out that motor WU has a different status at this time than verbal WU. The massed group shows several instances of reminiscence from the final prerest trial to the first postrest trial, but the steep, initial rise in each postrest segment is taken to be WU resulting from a decremental process opposing the gain over rest. Adams measured WU as the difference between the first postrest trial and the score on the trial at the peak of the rise before the decremental segment begins. For the distributed group, however, the method of WU measurement can be the same or it can be measured as a decrement from the last prerest trial to the first postrest trial because reminiscence is absent (Barch, 1954; Reynolds & Adams, 1954). Being able to measure it as an actual decrement is somewhat more precise because it does not involve judgments of the termination point of the WU segment. Digman (1959) replicated Adams' study in most of its aspects and obtained the same trends.

#### EXPLANATORY HYPOTHESES

##### *Set*

WU must be defined in terms of operations independent of those for a one-factor forgetting interpretation which, for the interference hypothesis of forgetting, would be in terms of responses conflicting with goal responses and causing extinction of them. Considering WU as a performance level below that expected at the beginning of a postrest practice session, it is just as meaningful to regard it as a simple one-factor forgetting loss for the goal response, with WU being a completely superfluous notion. With the exception of Doré and Hilgard (1938) and Hilgard and Smith (1942), pre-World War II investigators demonstrated a lack of methodological caution by simply assuming WU as a phenomenon separate from one-factor forgetting. The Iowa studies of psychomotor interference in the postwar era (Lewis & McAllister, 1950; Lewis, McAllister, & Adams, 1951; Lewis, Shephard, & Adams, 1949; Lewis, Smith, & Mc-

Allister, 1952; Shephard, 1950; Shephard & Lewis, 1950), exhibited a similar conservatism by suggesting an interpretation of WU consistent with the one-factor interference theory of forgetting. They held that the learning of responses in a laboratory task involves the extinction of conflicting responses either from prior tasks learned in the laboratory or from extralaboratory tasks. When a rest period is introduced the extinguished responses spontaneously recover some of their strength and, when postrest practice is resumed, the increased strength of these responses results in heightened conflict with the goal responses and WU occurs. As postrest practice continues the conflicting responses are once again extinguished and WU dissipates. While these one-factor views are parsimonious, and are therefore desirable, the parsimony may be unwarranted. The "other functions" which Thorndike (1914) identified with WU hypothesizes that a one-factor interpretation is an insufficient explanation of WU, and Thorndike's early view is given a more explicit, and testable, expression in the set hypothesis of WU. In the postwar era Irion (1948) gave the first operationally independent statement of WU in terms of a set state of the subject, and this definition was distinct from a one-factor forgetting definition. The term "set" has a number of meanings in psychology (Gibson, 1941) but Irion provided a sufficiently sound operational definition of set within the WU context to provide testable predictions. Although inhibition hypotheses of WU have been attempted, and will be discussed, the set hypothesis has the most status and has been the framework for most of the systematic research on WU.

If set is to be objectively assessed for its utility in the scientific description of behavior, it must be defined in terms of manipulable environmental events, on the one hand, and objective measures of behavior on the other. Furthermore, its operational definition on the environmental side must be different from those defining other behavioral processes which, for our present purposes, is the differentiation of set and one-factor forgetting variables. The independence of defining operations is critical for testing a two-factor theory of forgetting even though forgetting and set loss exert highly similar effects on dependent response measures. Just as long as one-factor forgetting is defined by the retroactive and proactive inhibition paradigm with experimental extinction as the process, and WU is defined by other operations related to a different process such as set, they both can be retained for the description of behavior because they can be independently manipulated and measured. This would be the justification for a scientifically sound two-factor theory of forgetting. Irion's paper (1948) made the two-factor distinction for verbal learning and consequently has given a basis for the objective assessment of set as a determiner of behavior. The use of set with respect to motor behavior has not been grounded in definitions as clear as those for verbal behavior, but this will be discussed later.

Irion's conception of set has much in common with those of Bell (1942) and Ammons (1947a) for motor learning where set is considered to be an aggregate of postural and attentive adjustments which are positively related to performance of the goal response. Complex perform-

ance, such as the learning of a verbal list, involves more than the external goal stimuli which the experimenter has objectively defined and controls, and to which the subject links the goal response measured by the experimenter. In addition, various secondary responses are learned, such as the orientation patterns for visual receptors, proper postural attitudes, and muscular tensions. These responses are secondary mainly in the sense of not being directly measured but the efficiency of the goal responding is intimately linked to them. Irion hypothesizes that these secondary responses, or set, are disturbed by the subject's activities between original learning and recall and this loss of set is the underlying cause for the steep slope of the initial segment of the relearning curve which is called WU. The disruption of set could operate to induce the decrement in retention in at least three ways: (a) failure of the receptors to adequately receive the goal stimuli, (b) mechanical inefficiency for optimum goal responding because the subject does not have the proper posture or muscular tension patterns, and (c) change in the internal stimulation which is part of the stimulus complex to which goal responses are conditioned (Guthrie, 1952). This third possible cause of WU could be a function of one or both of the first two because if the secondary responses are disturbed, then their patterns of response-produced stimulation change and the performance level of the goal responses conditioned to these internal cues is reduced.

While set is disrupted by activities during the rest interval, and thus is a kind of interference theory, the hypothesis is distinguished from the one-factor interference theory of forget-

ting by emphasizing the role of nongoal, secondary responses and, importantly, by specifying that these secondary responses are a function of operations different from those defining the strength of goal responses. The interference theory is concerned with practice variables which strengthen goal S-R sequences and increase their resistance to forgetting, and interfering S-R sequences which weaken goal S-R sequences by experimental extinction. Set, on the other hand, is strengthened by performance of S-R sequences that are neutral with respect to S-R goal sequences and which overcome WU by restrengthening secondary responses—not the strength of goal S-R sequences. While it is true that practice of goal responses appears to strengthen set, as the elimination of WU in relearning testifies, this is only because the goal responses are enmeshed in a matrix of secondary responses and their practice is concurrently accompanied by the practice of secondary responses. However, the task elements which define the learning problem for secondary responses can be embodied in a separate neutral task and can be used to strengthen set independently of goal response practicing. The weakening of set during the retention interval is also presumed to be by interfering activities neutral to goal responses, but their characteristics are unspecified at this time. It might be presumed, for example, that general body movements would disrupt the particular postures and muscular tension patterns acquired in the criterion task.

Of course there may be nothing to the set hypothesis because all retention loss could be one-factor forgetting in terms of direct effects on goal responses. But even given the general

terms within which the hypothesis is stated, the scientific criteria are broadly met for verifying that a portion of the retention loss can be ascribed to something other than one-factor forgetting, and it should be possible to find neutral tasks whose performance in a retention interval would reinstate set and abolish WU but would not yield habit strength increments for goal responses. Furthermore, if set is a determiner of performance as Irion says, practice on a neutral task should enhance performance on a criterion task before original learning by strengthening advantageous secondary responses. Effects on recall and original learning will be treated separately in the sections that follow.

#### *Verbal Behavior*

*Recall.* Taking cue from Ward's experiment (Ward, 1937) where the subject's association of colors during rest benefited verbal recall, Irion (1949b) tested the set hypothesis by using one trial of a neutral color-naming task as a warming-up activity just before the recall of paired adjectives after 24 hours. The subject was not required to memorize colors but only to name them as they appeared in the window of a memory drum. Color-naming, then, did not in any way involve practice of goal responses but did serve to reorient the subject to the rhythm of responding and direct his visual attending, posture, and physical adjustments in a manner very similar to that required in the criterion task and should function to restore the subject's set to respond. Irion found that a rest control group which had conventional recall after the 24-hour interval displayed a significant performance loss, but the color-naming group on

the first recall trial was significantly superior to the rest control group, and not different from a no-rest control group. This is in good accord with the set hypothesis. In re-establishing performance in the relearning trials, Irion found the one trial of color-naming essentially equivalent to one trial of practice on the criterion task. A related study reported in the same paper demonstrated that first trial recall was a decreasing function of the length of the rest interval up to 24 hours and that the slopes of the relearning curves were a function of the length of the rest interval. Irion interpreted this experiment as being in accord with the set hypothesis and the definition of WU in terms of the slope of the postrest performance curve. Since his color-naming experiment demonstrated that retention loss occurring after 24 hours could be eliminated by one trial of warming-up activity, it seems safe to assume that this decreasing recall function shows increasing WU and loss of set over interpolated time.

Irion and Wham (1951) tested an implication of the set hypothesis that WU should be a decreasing function of the amount of set-reinstating activity. The criterion task was serial role learning of nonsense syllables and the warming-up activity was recitation of three-place numbers. The retention interval was 35 minutes. Warming-up had a significant effect on the first recall trial, with performance level being a positive function of the number-naming trials. And, rate of increase of the initial WU segment of the relearning curves tended to be inversely related to the amount of warming-up. This study extends Irion's earlier work and represents good support for the set version of WU.

One interpretation that could be

given the positive effects of warming-up activities on verbal recall is that warming-up actually amounts to the strengthening of generalized techniques or modes of attack (learning how-to learn) and might be expected to appear as a higher level of goal responding in the initial recall trials. According to the interference theory of forgetting we would expect these general responses to be relatively independent of interpolated time because of the low probability of interfering responses occurring in casual experience. Set might be considered a more labile phenomenon, particularly if it is associated with muscular and postural patterns which could be disturbed by ordinary bodily movements that occur plentifully even during a relatively brief period of unstructured rest. Hartley (1948) considered this problem with an experiment which manipulated the temporal point of the warming-up activity. Using paired-associate adjectives as the learning task and color-naming as the warming-up activity, Hartley hypothesized that the set hypothesis would require color-naming to be effective in dissipating WU after a 24-hour rest only if given just prior to recall. But, if generalized habits were learned, then the expectation would be that color-naming immediately after original learning would beneficially influence recall. Hartley's data confirmed the set hypothesis. The group which had color-naming just prior to recall had the same performance level as a no-rest control group, but the group which had color-naming immediately after original learning had the same level of recall as a control group which had simple rest for 24 hours. Hunter (1955) demonstrated a similar effect of the time interval between warming-up activities and recall.

*Original learning.* If set is a general determiner of verbal behavior it is a reasonable expectation that warming-up activities just prior to original learning should increase performance level just as it does for recall. Heron (1928) performed the first experiment along these lines. Each subject learned two different lists of nonsense syllables a day on each of three different days. Heron found that the second list of each day was learned in fewer trials than the first list and he interpreted the positive transfer from the first to the second list as a temporary warming-up effect which dissipated over the time period between practice sessions. Thune (1951) refined the earlier work of Heron. Three different lists of paired adjectives were administered on each of five different days. While there was a slow learning-how-to learn effect as evidenced by the gradual upward trend of overall performance, the within-session gains from the first to third list each day were dramatic. Moreover, these gains largely disappeared during the rest interval between sessions. As Heron, Thune interpreted the gains within a session as a warming-up effect, with practice of each list exercising a facilitative effect on the learning of subsequent lists during a session because of set being re-established. Being labile, set is lost between sessions and thus there is relatively poor performance again at the beginning of each session.

Another experiment by Thune (1950) showed the positive effects of two kinds of warming-up activity on the original learning of a criterion list of paired nouns. Using a list of paired adjectives as the neutral warming-up activity, Thune held the total amount of practice on it constant but gave part of the practice on the first day and part of it on the second day just prior to learning the

criterion list. The experimental variable was the proportion of the total practice for the warming-up list given on the second day. Positive transfer to the criterion list was found to be an increasing function of this proportion, and the benefits were located in the early trials where the primary growth of set presumably takes place. In another experiment reported in the same paper, Thune found that 10 trials of guessing the color that would appear next under the shutter of the memory drum had about the same positive transfer effects as 10 trials of the neutral list of nouns if both were presented just before practice of the criterion list. By using appropriate control groups, Thune further showed that 10 trials of color-guessing administered 24 hours before the test list had no effect and he was drawn to the same conclusions as Hartley (1948) that the temporal point of warming-up is critical, and neutral warming-up activities function to strengthen labile set factors rather than stable habits. All of these findings of Thune's are in agreement with the set hypothesis. Hamilton (1950) performed an experiment closely coordinated with the Thune investigation in which the time interval between the warming-up list and the test list was varied. Performance on the criterion list was an exponential decay function of the retention interval and was interpreted as the curve for the loss of unstable set factors.

#### *Motor Behavior*

There is no intrinsic reason why the set hypothesis should not apply to both motor behavior and verbal behavior and, indeed, this would be a desirable generality. WU-like characteristics are commonly found in post-rest motor performance curves and set is a promising candidate for explanation, but the problem for motor

behavior lies in differentiating set and one-factor forgetting as determiners of the WU segment of the curve. The means of doing this for verbal learning and recall mainly was to show positive effects of goal behavior from performance on neutral tasks that could not reasonably be thought as strengthening goal responses and therefore must be enhancing something else—set. The difficulty for motor behavior at this time is that these neutral tasks have not yet been discovered, if they exist at all, and so the WU-like effects can just as well be explained by a one-factor forgetting theory. Certainly this difficulty does not negate the set hypothesis, and actually it may only be a temporary impasse, but it does urge restraint upon us.

Ammons' research on WU illustrates the problem (Ammons, 1947a, 1947b, 1950). The basis of his work is a miniature theoretical system for rotary pursuit performance (Ammons, 1947a) which leans heavily on Hull's theoretical system for explanatory constructs (Hull, 1943) but has the added feature of set as a construct term to explain the WU segment. Ammons specifies set and WU as inversely related, and relates them to the independent variables of number of prerest practice sessions, total amount of prior practice, duration of practice sessions, duration of interpolated rest, and the amount of practice that has occurred in the present, ongoing practice period. In subsequent papers on empirical research (Ammons 1947b, 1950) Ammons reports a number of relationships between WU and these independent variables but, since the one-factor interference theory of forgetting would also relate retention to rest and to practice variables defining habit strength in the task, there is no foundation for the assumption that these

independent variables are manipulating WU through changes in set. Irion (1949a) performed similar research with the Rotary Pursuit Test and related WU to the amount of pre-rest practice and the duration of rest. His paper reflects the same methodological problem.

Efforts to locate a neutral task which would influence WU in the recall of motor performance, as well as the level of original learning, have met with failure. The most thorough experimental attempt was by Ammons (1951) using the Rotary Pursuit Test. His subjects were administered initial practice, rest, and a postrest practice session. Set-reinstating activities were either watching the disk, blindfolded manual performance of the rotary motion by holding a small rivet set in the rotor plate, or imaginary practice where the subject was merely to think about practicing. These activities were administered either before initial practice, before the postrest practice session, or before both practice periods. No effects were observed from any of the experimental treatments. Walker, DeSoto, and Shelly (1957) performed a bilateral transfer experiment on the Rotary Pursuit Test. Original practice was with one hand and, following rest, practice was resumed with the other hand. One of the experimental conditions was to have one trial of practice just before the postrest session with the pre-rest practice hand to see if it could have a warming-up effect on the transfer hand. WU was found in performance on the transfer hand but it was unaffected by the warming-up procedure and they concluded that WU must be quite specific to an effector. Hamilton and Mola (1953) used a finger maze and evaluated the effect of practice on five different mazes on

performance in a criterion maze. They used an experimental design similar to Hartley (1948) and Thune (1950) and gave practice on the five warming-up mazes either 24 hours or immediately preceding the test maze. Positive transfer to the criterion maze was found but it was about the same for both warm-up groups and the authors concluded that practice on the five mazes exerted only a general practice effect and had no warming-up properties. A small encouraging sign to counter this negative evidence is found in a study by Adams (1955) on sources of work inhibition in complex motor performance. The Rotary Pursuit Test was used and during an intersession period one group was required to observe a partner's performance and press a button every time he judged him to be on target. This activity produced work inhibition but it also tended to result in less WU than found for control groups. The reduced WU was a secondary finding in a study on another topic but it is a lead on a likely set-reinstating activity for motor performance.

#### *Negative Evidence*

Ordinarily the amount of evidence which has been cited in support of the set hypothesis for verbal learning would be sufficient to give a good measure of security to the two-factor theory in psychology, but unfortunately there is a disconcerting number of negative findings. In a careful effort to replicate Irion's verbal learning study (1949b), Rockway and Duncan (1952) were unable to reproduce Irion's results and show an effect of color-naming on recall. Similarly, Withey, Buxton, and Elkin (1949) and Hovland and Kurtz (1951) failed to show an influence of

color-naming on verbal recall. Underwood (1952), studying the serial learning of nonsense syllables, was unable to demonstrate that the warming-up activity of number-naming prior to recall produced an effect on WU after 24 hours of rest. Underwood said that this did not necessarily contradict previous findings by Irion and others because earlier studies did not use the same subjects in several experimental conditions as he had done. Dinner and Duncan (1959) hypothesized that the unreliability of the effects of color-naming on verbal recall might be a function of degree of original learning. Using a low, medium, and high degree of original learning of paired adjectives, they found that color-naming influenced recall only when level of original learning was high. They concluded that Irion's positive use of color-naming (Irion, 1949b) based on a medium degree of original learning should be considered sampling error and discounted. However, this judgment should be regarded with caution because it does not consider the works of Hartley (1948) and Irion and Wham (1951) who all obtained positive effects of warming-up activities on verbal recall when low or intermediate levels of original learning were used. The Dinner and Duncan investigation makes an original contribution in showing an effect of the amount of original learning, but it cannot be reconciled at this time with verbal learning studies which effectively used warming-up activities to enhance recall.

Another disturbing consideration for understanding WU and the set hypothesis is that there are tasks where performance in the initial postrest trials does not show WU. The inverted alphabet printing task has enjoyed moderate popularity for the

study of work inhibition and the data never show WU (Archer, 1954; Eysenck, 1956; Kimble, 1949; Wasserman, 1951). Silver (1952) used the inverted alphabet printing task to investigate the interaction of warm-up activity on WU and work inhibition, but since he used performance on the first postrest trial for his comparisons there is no reason to believe that he was manipulating WU or, indeed whether his data even displayed WU. Other investigators using inverted alphabet printing failed to show WU, and it is unlikely that WU as it has been defined in terms of a rapid increase in performance on the initial postrest trials was even present in Silver's data. Bilodeau (1952a, 1952b), investigated work inhibition by manipulating the physical load required to turn a manual crank, and found no WU in his data. Doten (1955), in a study of interference, used a task where the subject was presented the printed names of colors, but where the color of the letters in the name was different than indicated by the name. The word "Red" might be printed in blue, for example. The task of the subject in original learning was to respond by stating the actual color of the lettering, and no WU was indicated. The initial segments of performance curves on each day had an immediate decrease in speed of responding, not the rapid increase which is characteristic of WU segments. Tasks such as these raise serious definitional problems for set, or any other WU hypothesis for that matter. We cannot expect any explanatory hypothesis to enjoy a good degree of success until the tasks in which WU occurs have been established. Ideally the set hypothesis should contain statements relating WU and task characteristics.

*Inhibition*

The principal advocate of an inhibition hypothesis is Eysenck (1956) who interprets WU as mainly being attributable to the extinction of Hull's conditioned inhibition (Hull, 1943). Eysenck does not completely deny the set hypothesis, but rather considers loss of set a lesser contributor to WU, with the extinction of conditioned inhibition being the primary reason for the trend of the initial postrest segment. It will be recalled that Hull has a two-factor theory of inhibition. One construct,  $I_R$ , is an increasing function of the number of responses and amount of physical work, and a decreasing function of the rest interval. Moreover,  $I_R$  has drive properties and its dissipation is regarded as drive reduction. Since drive reduction in Hull's system is the basis of reinforcement, an increment of habit strength for the on-going response is accrued whenever  $I_R$  dissipates. Because the subject is resting when  $I_R$  is dissipating, it is theorized that a resting response is strengthened which is antagonistic to the goal response. The habit construct for the resting response is the second inhibitory factor,  $sI_R$ . These two types of inhibition summate and subtract from the excitatory potential ( $sE_R$ ) for the goal response to yield effective excitatory potential ( $gE_R$ ) which is the primary determiner of overt performance level. The massed group in Figure 1 can be used to illustrate Eysenck's application of Hull's inhibition theory to WU. In the first session the subject responds continuously and  $I_R$  accrues. Then, over rest,  $I_R$  dissipates and an increment of  $sI_R$  develops. The failure of performance on the first postrest trial to reminisce to the level of the distributed group is taken as evidence for the presence of  $sI_R$ .

When the subject begins practice in the second session the goal response is now being reinforced and the non-reinforced resting response undergoes experimental extinction. This period of extinction of the resting response is revealed as the WU segment, according to Eysenck. One immediate prediction from Eysenck's hypothesis is that little WU should be found under well-spaced practice conditions where a negligible amount of  $I_R$  is generated on each trial, and thus negligible  $sI_R$ . Eysenck tested this deduction using the Rotary Pursuit Test and, in accord with his prediction, found WU under conditions of massed practice but not distributed practice. His findings and conclusions are tenuous however, because of the instances in the experimental literature showing WU under conditions of widely distributed practice on the Rotary Pursuit Test. Figure 1 is a good example (Adams, 1952b). Other examples are Ammons (1950), Denny, Frisbey, and Weaver (1955), Digman (1959), Kimble and Shatel (1952), and Jahnke and Duncan (1956). There is no immediate explanation for Eysenck's unusual finding, but WU under conditions of well-distributed practice is a commonplace finding and suggests that Eysenck's hypothesis cannot be taken seriously.

Adams (1952b) entertained a different inhibition hypothesis. Observing that much of the evidence for WU came from studies on the Rotary Pursuit Test under conditions of massed practice, he deduced the characteristics of a postrest performance curve from a negatively accelerated growth of reaction potential with trials and an ogival function for the accrual of work inhibition when practice is massed. It was predicted that WU should not appear under

conditions of distributed practice. The occurrence of clearcut WU when training was widely distributed (Figure 1) led to rejection of the hypothesis.

At present it must be concluded that no convincing evidence exists in support of an inhibition explanation of WU.

#### WARM-UP IN ANIMALS

The main concern over WU has been with human subjects, but it is noteworthy that the phenomenon also has been observed in animals. The following studies are not meant to represent an exhaustive search of the literature on animal behavior, but rather are intended to show the ubiquity of WU-like effects and that its characteristics are not found only in human response records. Schlosberg (1934, 1936) interprets as WU the failure of occurrence of a well-learned conditioned response in the white rat on the first few trials of a learning session. Ellson (1938), in studying extinction of a bar-pressing habit in the rat, found rate of response slower in the first fifth of the extinction trials than in the second fifth. He interpreted this as WU and explained it in terms of Guthrie's theory which holds that the stimuli to which a response is learned include internal stimuli resulting from posture, movement, etc. Later responses are partly conditioned to the response-produced stimuli of earlier responses and we would expect that later responses in a series would have greater strength because of the presence of the response-produced stimuli to which they are conditioned. In the latter part of extinction the effects of nonreward overcome this trend and the performance level then decreases systematically. Finger (1942) used rats in a straight runway situation

and found WU revealed when an extinction series was administered after 24 hours. The second extinction trial actually had performance superior to that of the first extinction trial—a finding contrary to expectation for an extinction series. Finger's finding for extinction is quite similar to Ellson's. Verplanck (1942) reported WU for rats in a simple running task. Like many investigators of human behavior, these animal researchers freely labeled decrements in initial postrest segments of performance records as WU although the decrements could just as well, and more economically, have been explained by the one-factor forgetting hypothesis.

#### DISCUSSIONS AND CONCLUSIONS

Virtually all support for the two-factor theory of forgetting is embodied in experiments which have demonstrated that WU is reduced or eliminated by repetition of responses that orient the subject to the general task demands (e.g., color-naming) but which do not involve direct practice of goal responses. By themselves these experiments might be sufficient to establish set as a second factor necessary for the explanation of retention loss, but the studies where set-reinstating activities have failed to influence recall in both motor and verbal tasks, and the tasks where no WU whatsoever has been found, leave the second factor in doubt. There is not sufficient evidence to reject the set hypothesis but neither are there grounds for firmly retaining it. Certainly it is the most tenable of all hypotheses advanced, but a great deal of careful research seems required before the set hypothesis, and thus the two-factor theory of forgetting, can be accepted or rejected with confidence.

It is unlikely that a decision ever

will be made about the set hypothesis unless it receives a more thorough testing than it has in the past. The set-reinstating experiments are very broadly derived from the hypothesis and have not been a test of the more explicit implications of set. By viewing the set hypothesis in its more detailed aspects, and in attempting to develop specific experiments and measures to empirically verify these details, it should be possible not only to clarify the status of the set hypothesis but also to determine why, for example, some tasks display WU and others do not. For example, Irion (1948), Ammons (1947a), and Bell (1942) all contend that the acquisition of set can be the learning of beneficial postures and muscular tensions that facilitate the occurrence of goal response sequences. Rest period activities disturb the favorable set and the WU segment of a post-rest performance curve represents the reacquisition of these favorable bodily attitudes. If there is anything to this version of the set hypothesis, it would seem fruitful to explore the characteristics of bodily tensions by direct measurement and then relate it to changes in performance of the goal response. Davis and his associates have performed a number of studies (e.g., Davis, 1940, 1956) showing the relationship between the characteristics of overt responding and muscular tensions as revealed by electromyographic measurement techniques. Davis (1956) does not believe that the muscular substrata and the overt goal responses need be conceptualized as fundamentally different. A state of tension in skeletal muscle is the same as any other muscular contraction, i.e., it is a response configuration. Davis (1956) says:

Muscular tensions would then be themselves

responses to stimuli, many being small responses, detectable only with instruments, but with no firm boundary between them and the larger muscular activities associated with movement (p. 2).

Davis' work is suggestive for the set hypothesis because it strongly hints that the pattern of electromyographic measures of muscular tension during the WU segment of the post-rest performance curve would have levels and patterns of muscular tension different from final prerest performance, and these levels and patterns will have shifted in the direction associated with poorer performance. Moreover, the reacquisition of prerest values and patterns of muscular tensions should parallel the trend of the WU segment. Furthermore, and importantly, it suggests that neutral set-reinstating activities will produce electromyographic changes signifying that the favorable muscular tensions existing at final prerest performance are being re-established.

There are difficulties in operationally distinguishing between an electromyographically-verified muscle tension version of the set hypothesis and Irion's alternative Guthrian hypothesis that loss of set is disturbance of internal stimuli to which goal responses are partly conditioned. If we have changes in the muscle tension secondary responses and this in turn, results in a lower level for goal responses, we cannot be sure that the lower level is due to quasi-mechanical considerations where muscular tensions underlie useful postures and bodily attitudes, or whether it is due to changes in the population of stimuli to which the goal responses have been conditioned. Despite the potential difficulties of interpreting the primary effects of muscle tension secondary responses, on performance of the goal responses, it would be a fun-

damental finding to show a systematic covariation of electromyographic measures and WU phenomena. The evanescent quality of set could benefit from a diversity of approaches at this time to provide clues for a reconciliation of inconsistencies among the various experimental findings.

The delineation of set and its role in retention will sharpen our understanding of the retention loss problem and will improve our efforts to predict and control it. Underwood (1957) has shown that our frequent use of the same subjects in several laboratory experiments has led us to greatly overestimate the retention loss for verbal responses because the experimenter was unwittingly contaminating his retention scores with proactive inhibition effects. But even given this downward revision of retention loss, we are still faced with

showing the proportion of it attributable to interference with goal responses and the part assignable to loss of set. Irion's experiment (1949b), for example, showed that one trial of color-naming almost completely eliminated the verbal retention loss and therefore all of the loss could be described in terms of change in set. This suggests that if the two-factor theory eventually becomes better established in fact the paradigm of retention studies will have to include groups whose performance of set-reinstating activities will allow a parsing of set and interference components. Interference with goal responses may be a smaller contributor to retention loss than we now surmise. The first research need however, is a more incisive laboratory attack on the validity of set and its underlying nature.

## REFERENCES

ADAMS, J. A. The influence of the time interval after interpolated activity on psychomotor performance. *USAF Hum. Resour. Res. Cen. res. Bull.*, 1952, No. 52-11, (a)

ADAMS, J. A. Warm-up decrement in performance on the pursuit-rotor. *Amer. J. Psychol.*, 1952, **65**, 404-414. (b)

ADAMS, J. A. A source of decrement in psychomotor performance. *J. exp. Psychol.*, 1955, **49**, 390-394.

AMMONS, R. B. Acquisition of motor skill: I. Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263-281. (a)

AMMONS, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393-411. (b)

AMMONS, R. B. Acquisition of motor skill: III. Effects of initially distributed practice on rotary pursuit performance. *J. exp. Psychol.*, 1950, **40**, 777-787.

AMMONS, R. B. Effects of prepractice activities on rotary pursuit performance. *J. exp. Psychol.*, 1951, **41**, 187-191.

ARAI, T. Mental fatigue. *Teach. Coll. Contrib. Educ.*, 1912, No. 54.

ARCHER, J. E. Postrest performance in motor learning as a function of prerest degree of distribution of practice. *J. exp. Psychol.*, 1954, **47**, 47-51.

BARCH, A. M. Warm-up in massed and distributed pursuit rotor performance. *J. exp. Psychol.*, 1954, **47**, 357-361.

BELL, H. M. Rest pauses in motor learning as related to Snoddy's hypothesis of mental growth. *Psychol. Monogr.*, 1942, **54**(1, Whole No. 243).

BILODEAU, E. A. Decrement and recovery from decrements in a simple work task with variation in force requirements at different stages of practice. *J. exp. Psychol.*, 1952, **44**, 96-100. (a)

BILODEAU, E. A. Massing and spacing phenomena as a function of prolonged and extended practice. *J. exp. Psychol.*, 1952, **44**, 108-113. (b)

BRIGGS, G. E. Acquisition, extinction, and recovery function in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285-293.

BRIGGS, G. E. Retroactive inhibition as a function of the degree of original and interpolated learning. *J. exp. Psychol.*, 1957, **53**, 60-67.

BUGELSKI, B. R., & CADWALLADER, T. C. A

reappraisal of the transfer and retroaction surface. *J. exp. Psychol.*, 1956, 52, 360-366.

DAVIS, R. C. Set and muscular tension. *Indiana U. Publ. Sci. Ser.*, 1940, No. 10.

DAVIS, R. C. Electromyographic factors in aircraft control: The relation of muscular tension to performance. *USAFA Sch. Aviat. Med. Rep.*, 1956, No. 55-122.

DENNY, R. M., FRISBEY, N., & WEAVER, J., JR. Rotary pursuit performance under alternate conditions of distributed and massed practice. *J. exp. Psychol.*, 1955, 49, 48-54.

DIGMAN, J. M. Growth of a motor skill as a function of distribution of practice. *J. exp. Psychol.*, 1959, 57, 310-316.

DINNER, JUDITH E., & DUNCAN, C. P. Warm-up in retention as a function of degree of verbal learning. *J. exp. Psychol.*, 1959, 57, 257-261.

DORÉ, L. R., & HILGARD, E. R. Spaced practice as a test of Snoddy's two processes in mental growth. *J. exp. Psychol.*, 1938, 23, 359-374.

DOTEN, G. W. The effects of rest periods on interference of a well-established habit. *J. exp. Psychol.*, 1955, 49, 401-406.

ELLISON, D. G. Quantitative studies of the interaction of simple habits: I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, 23, 339-358.

EYSENCK, H. J. "Warm-up" in pursuit rotor learning as a function of the extinction of conditioned inhibition. *Acta psychol. Amst.*, 1956, 12, 349-370.

FINGER, F. W. Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, 1942, 31, 120-133.

GIBSON, J. J. A critical review of the concept of set in contemporary experimental psychology. *Psychol. Bull.*, 1941, 38, 781-817.

GUTHRIE, E. R. *The psychology of learning*. (Rev. ed.) New York: Harper, 1952.

HAMILTON, C. E. The relationship between length of interval separating two learning tasks and performance on the second task. *J. exp. Psychol.*, 1950, 40, 613-621.

HAMILTON, C. E., & MOLA, W. R. Warm-up effect in human maze learning. *J. exp. Psychol.*, 1953, 45, 437-441.

HARTLEY, T. C. Retention as a function of the temporal position of an interpolated warming-up task. Unpublished MA thesis, University of Illinois, 1948.

HERON, W. T. The warming-up effect in learning nonsense syllables. *J. genet. Psychol.*, 1928, 35, 219-228.

HILGARD, E. R., & SMITH, M. B. Distributed practice in motor learning: Score changes within and between daily sessions. *J. exp. Psychol.*, 1942, 30, 136-146.

HOVLAND, C. I., & KURTZ, K. H. Experimental studies in rote-learning theory: IX. Influence of work-decrement factors on verbal learning. *J. exp. Psychol.*, 1951, 42, 265-272.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

HUNTER, I. A. The warming-up effect in recall performance. *Quart. J. exp. Psychol.*, 1955, 7, 166-175.

IRION, A. L. The relation of "set" to retention. *Psychol. Rev.*, 1948, 55, 336-341.

IRION, A. L. Reminiscence in pursuit-rotor learning as a function of length of rest and of amount of pre-rest practice. *J. exp. Psychol.*, 1949, 39, 492-499. (a)

IRION, A. L. Retention and warming-up effects in paired associate learning. *J. exp. Psychol.*, 1949, 39, 669-675. (b)

IRION, A. L., & WHAM, DOROTHY S. Recovery from retention loss as a function of amount of pre-recall warming-up. *J. exp. Psychol.*, 1951, 41, 242-246.

JAHNKE, J. C., & DUNCAN, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, 52, 273-282.

KIMBLE, G. A. An experimental test of two-factor theory of inhibition. *J. exp. Psychol.*, 1949, 39, 15-23.

KIMBLE, G. A., & SHATEL, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, 44, 355-359.

LEWIS, D., & McALLISTER, DOROTHY E. An investigation of individual susceptibility to interference. *USN Spec. Dev. Cent. tech. Rep.*, 1950, No. 938-1-10.

LEWIS, D., McALLISTER, DOROTHY E., & ADAMS, J. A. Facilitation and interference in performance on the modified Mashburn apparatus: I. The effects of varying the amount of original learning. *J. exp. Psychol.*, 1951, 41, 247-260.

LEWIS, D., SHEPHARD, A. H., & ADAMS, J. A. Evidences of associative interferences in psychomotor performance. *Science*, 1949, 110, 271-273.

LEWIS, D., SMITH, P. N., & McALLISTER, DOROTHY E. Retroactive facilitation and interference in performance on the Modified Two-Hand Coordinator. *J. exp. Psychol.*, 1952, 44, 44-50.

MELTON, A. W. (Ed.) *Apparatus tests*. (AAF Aviat. Psychol. Program res. Rep. No. 4) Washington, D. C.: United States Government Printing Office, 1947.

MOSKO, A. *Fatigue*. (Trans. by M. Drummond) New York: Putnam, 1906.

OSGOOD, C. E. The similarity paradox in human learning. *Psychol. Rev.*, 1949, 56, 132-143.

OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford, 1953.

REYNOLDS, B., & ADAMS, J. A. Psychomotor performance as a function of initial level of ability. *Amer. J. Psychol.*, 1954, 67, 268-277.

ROBINSON, E. S. Work of the integrated organism. In C. Murchison (Ed.), *Handbook of general experimental psychology*, 1934.

ROBINSON, E. S., & HERON, W. T. The warming-up effect. *J. exp. Psychol.*, 1924, 7, 81-97.

ROCKWAY, M. R., & DUNCAN, C. P. Pre-call warming-up in verbal retention. *J. exp. Psychol.*, 1952, 43, 305-312.

SCHLOSBERG, H. Conditioned responses in the white rat. *J. genet. Psychol.*, 1934, 45, 303-335.

SCHLOSBERG, H. Conditioned responses in the white rat: II. Conditioned responses based upon shock to the foreleg. *J. genet. Psychol.*, 1936, 49, 107-138.

SHEPHERD, A. H. Losses of skill in performing the standard Mashburn task arising from different levels of learning on the reversed task. *USN Spec. Dev. Cent. tech. Rep.*, 1950, No. 938-1-9.

SHEPHERD, A. H., & LEWIS, D. Prior learning as a factor in shaping performance curves. *USN Spec. Dev. Cent. tech. Rep.*, 1950, No. 938-1-4.

SILVER, R. J. Effect of amount and distribution of warming-up activity on retention in motor learning. *J. exp. Psychol.*, 1952, 44, 88-95.

SNODDY, G. S. *Evidence for two opposed processes in mental growth*. Lancaster: Science, 1935.

STEINBERG, HANNAH, & SUMMERFIELD, A. Influence of a depressant drug on acquisition in rote learning. *Quart. J. exp. Psychol.*, 1957, 9, 138-145.

SUMMERFIELD, A., & STEINBERG, HANNAH. Reducing interference in forgetting. *Quart. J. exp. Psychol.*, 1957, 9, 146-154.

SUMMERFIELD, A., & STEINBERG, HANNAH. Using drugs to alter memory experimentally in man. In P. B. Bradley, P. Deniker, & C. Radouco-Thomas (Eds.), *Neuro-psychopharmacology*. Houston: Elsevier, 1959. Pp. 481-483.

THORNDIKE, E. L. *Educational psychology*. Vol. 3. New York: Teachers Coll., Columbia Univer., 1914.

THUNE, L. E. The effects of different types of preliminary activities on subsequent learning of paired-associate learning. *J. exp. Psychol.*, 1950, 40, 423-438.

THUNE, L. E. Warm-up effect as a function of level of practice in verbal learning. *J. exp. Psychol.*, 1951, 42, 250-256.

UNDERWOOD, B. J. Retroactive and proactive inhibition after five and forty-eight hours. *J. exp. Psychol.*, 1948, 38, 29-38. (a)

UNDERWOOD, B. J. "Spontaneous recovery" of verbal associations. *J. exp. Psychol.*, 1948, 38, 429-439. (b)

UNDERWOOD, B. J. Studies of distributed practice: VI. The influence of rest-interval activity in serial learning. *J. exp. Psychol.*, 1952, 43, 329-340.

UNDERWOOD, B. J. Interference and forgetting. *Psychol. Rev.*, 1957, 64, 49-60.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, 67, 73-95.

VERPLANCK, W. S. The development of discrimination in a simple locomotor habit. *J. exp. Psychol.*, 1942, 31, 441-464.

WALKER, L. C., DESOTO, C. B., & SHELLY, M. W. Rest and warm-up in bilateral transfer on a pursuit rotor task. *J. exp. Psychol.*, 1957, 53, 394-404.

WARD, L. B. Reminiscence and rote learning. *Psychol. Monogr.*, 1937, 49(4, Whole No. 220).

WASSERMAN, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, 42, 162-172.

WATSON, J. B. *Psychology from the standpoint of a behaviorist*. Philadelphia: Lippincott, 1919.

WELLS, F. L. Normal performance on the tapping test before and during practice with special reference to fatigue phenomenon. *Amer. J. Psychol.*, 1908, 19, 437-483.

WHITHEY, S., BUXTON, C. E., & ELKIN, A. Control of rest interval activities in serial verbal learning. *J. exp. Psychol.*, 1949, 39, 173-176.

(Received July 15, 1960)

## ON THE REFORMULATION OF INHIBITION IN HULL'S SYSTEM

ARTHUR R. JENSEN  
*University of California*

Among the least satisfactory elements of Hull's behavior system is his formulation of inhibition. As a result, there have been several attempts in recent years to reformulate Hull's theory with respect to the inhibition variables in the equation for effective reaction potential ( $s\bar{E}_R$ ). The present paper critically examines these reformulations in the light of relevant experimental evidence. The conclusions to which this examination leads are that these reformulations have not been an improvement over Hull and that this kind of reformulation itself is a futile approach to the problem of improving Hullian-type learning theory.

In all versions of his theory Hull (1943, 1951, 1952) formulated "effective reaction potential" ( $s\bar{E}_R$ ) as being essentially a function of "drive" ( $D$ ) and "habit strength" ( $sH_R$ ), related multiplicatively (i.e.,  $D \times sH_R$ ), minus "reactive inhibition" ( $I_R$ ) and "conditioned inhibition" ( $sI_R$ ), related additively (i.e.,  $I_R + sI_R$ ). Thus:

$$s\bar{E}_R = (D \times sH_R) - (I_R + sI_R)$$

Most of the attempts to reformulate Hull's equation have been the result of logical, or at times merely verbal, rather than empirical considerations. For example, Hilgard's (1956, p. 139) criticism is directed at the fact that Hull did not carry out the logical implications of his statement that  $I_R$  is a "negative drive state." As such,  $I_R$  logically should subtract from  $D$  (i.e.,  $D - I_R$ ) and, like  $D$ , should interact multiplicatively with habit strength (i.e.,

$I_R \times sH_R$ ). Hilgard also suggests that, since  $sI_R$  is a negative habit, it should interact multiplicatively with  $I_R$ . Thus, Hilgard's proposed reformulation of the equation for net reaction potential results in the following:

$$s\bar{E}_R = [(D - I_R) \times sH_R] - (I_R \times sI_R)$$

This new formulation seems to be more consistent with some of Hull's own statements about the nature of these intervening variables, but Hilgard avoids trouble by not attempting to relate this formulation to empirical findings.

Similarly, Iwahara (1957) carries Hull's characterization of  $I_R$  as a negative drive and  $sI_R$  as a negative habit to what may seem the logical conclusion in terms of the internal consistency of Hull's theory—that the relationship between drives and habits is always multiplicative and never additive. Iwahara then goes a step further to regard  $sI_R$  as a conditioned or secondary negative drive, with  $I_R$  being the primary negative drive. From this it follows that the product of  $I_R \times sI_R$  should subtract from positive drive,  $D$ , and should also multiply  $sH_R$ . Symbolically,

$$s\bar{E}_R = sH_R \times [D - (I_R \times sI_R)]$$

or, in expanded form,

$$s\bar{E}_R = (sH_R \times D) - (sH_R \times I_R \times sI_R)$$

Osgood (1953, p. 379) states that Hull need not have postulated  $sI_R$  at all, since it might have been derived from other postulates in the system. If  $sI_R$  is nothing other than

negative habit strength or the habit of not responding (reinforced by the dissipation of  $I_R$ ), it would seem logical to subtract  $sI_R$  directly from  $sH_R$ . This is the formulation Osgood has proposed (p. 349).

More recently, Jones (1958) has incorporated the foregoing suggestions in his revision of Hull's equation. The Jones version, which combines the properties of the other revisions (except Iwahara's  $sH_R \times sI_R$ ) and appears identical to Osgood's suggestion, is as follows:

$$s\bar{E}_R = (D - I_R) \times (sH_R - sI_R)$$

That this formulation is quite radically different from Hull's is even more obvious when Jones mathematically expands the equation, thus:

$$s\bar{E}_R = (D \times sH_R) - (I_R \times sH_R) \\ - (D \times sI_R) + (I_R \times sI_R)$$

Jones' formulation has been subscribed to by Eysenck and his co-workers in their attempt to utilize Hullian postulates in developing a theory of personality (Eysenck, 1957; Kendrick, 1958).

Another revision, rather casually suggested by Woodworth and Schlosberg (1954, p. 668), is that inhibition ( $I_R$  or  $sI_R$  or both?) should subtract from "incentive motivation" (Hull's  $K$ , a function of the amount of reinforcement). Presumably the total inhibitory potential  $I_R$  (the sum of  $I_R + sI_R$ ) subtracts from  $K$ , though this point is not clear in the Woodworth and Schlosberg discussion. Their suggestion might be expressed symbolically as follows:

$$s\bar{E}_R = (K - I_R - sI_R) \times D \times sH_R$$

The most carefully formulated and empirically anchored modifications of Hull's theory have been those of Spence (1956). His changes in the

inhibition part of the theory are of a fundamentally different nature than the other revisions. He has more or less wiped the slate clean and started anew by redefining inhibition and the independent variables of which it is a function. Spence's extinctive inhibition ( $I_n$ ) is not a function of the amount of effort or rate of responding, as is Hull's  $I_R$ , but is a function only of the number of nonreinforced responses. There is also an oscillatory inhibition ( $I_o$ ), which is the same as Hull's concept of oscillation ( $sO_R$ ). The inhibition due to delay of reward ( $I_t$ ) is essentially the same as  $I_n$ . The basis of this inhibition is assumed to be the competing responses that are established during the delay period or during extinction. The molar concepts of  $I_t$  or  $I_n$  simply represent the quantitative effects of these competing responses. Spence's inhibition does not interact with other intervening variables but only subtracts from the reaction potential. In this last respect his formulation is essentially no different from Hull's. It might be asked why  $D$ , if it is regarded as an energizer of all responses in the organism's repertoire, should not interact with inhibition as Spence conceives of it, that is, as consisting of interfering or competing responses. In this respect Spence's theory of extinction is not unlike Guthrie's.

With the exception of Spence, these attempts to reformulate Hull raise a number of crucial questions in common, some of which must be critically examined on the level of theory and methodology and others in terms of empirical evidence. First there are questions of a general theoretical nature which must be considered in relation to any attempt to criticize or reformulate Hull's theory.

1. Is the verbal formulation of

Hull's theory to be taken more seriously than the symbolic and quasi-quantitative formulations, or than the actual empirical relationships which formed the basis for Hull's postulates and which he has held up as examples of the relationships he wished his system to predict?

2. Does the algebraic manipulation of Hull's intervening variables make sense theoretically and psychologically? Are the functions representing their interrelationships "isomorphic" with the rules of simple algebra?

3. Can experiments be designed to determine the exact nature of the intervening variables?

Once one has decided to argue within the Hullian framework a number of questions arise from the attempts at reformulation, the answers to which must depend upon empirical findings.

1. Does  $sI_R$  subtract from  $sH_R$ ? Are  $sH_R$  and  $sI_R$  both basically the same phenomenon, one merely being positive and the other negative in effect, or do they represent basically different processes?

2. Is there any empirical evidence to support the following formulations?

a. The interaction of  $D \times sI_R$  (Jones, Osgood)

b.  $D - I_R$  (Hilgard, Jones, Osgood)

c. The interaction of  $sH_R \times I_R$  (Hilgard, Iwahara, Jones, Osgood)

d. The interaction of  $sH_R \times sI_R$  (Iwahara)

e. The interaction of  $I_R \times sI_R$ , which paradoxically represents an *addition* to reaction potential, the multiplication of two negative quantities making a positive (Hilgard, Iwahara, Jones, Osgood)

f.  $K - I_R$  (Woodworth & Schlosberg)

#### THE LIMITATIONS OF HULL'S THEORY

In offering his revision, Jones (1958) points out that the inhibition aspect of Hull's formula for reaction potential has been criticized by Koch (1954). Koch's criticisms, however, apply equally to Jones' revision as well as to all the others, with the possible exception of Spence. Koch points out that the intervening variables concerning inhibition in Hull's system, particularly  $sI_R$ , are not rigorously defined, are not clearly tied to experimental variables, and hence are indeterminate. Because of this, it is impossible to make rigorous experimental tests of Hull's formulations or of the alternative revisions. Cotton (1955) has shown that a literal interpretation of Hull's postulates leads to predictions that differ from the experimental data upon which Hull based the formulation of his postulates in the first place. In short, much of Hull's theory does not even predict the very facts it was expressly devised to predict. This is especially true with regard to the inhibition postulates. None of the revisions of Hull has improved this situation. They have merely rearranged in various ways the same indeterminate variables of Hull's formula for  $sE_R$ .

Hull's revisers have followed him in treating his intervening variables,  $D$ ,  $sH_R$ ,  $I_R$ ,  $sI_R$ , etc., as if they were real, independent quantities whose laws of interaction are isomorphic with the rules of arithmetic and algebra. As we shall see, the manipulation of these hypothetical variables in such fashion can at times lead to absurdity. Hull's intervening variables are only intervening variables in the sense which MacCorquodale and Meehl (1948) have assigned to that term, and are defined only in terms of the independent and dependent

variables to which they are tied. The danger arises when Hull's revisers mathematically manipulate the intervening variables without regard for the defining experimental variables which are actually all that give any meaning to the intervening variables. Of course, one of the purported virtues of intervening variables is that they can be mathematically manipulated as independent entities. But once the intervening variable has been properly defined, the question arises as to the nature of the mathematical operations that can suitably be applied to it. It is highly doubtful if the exclusive use of linear algebra by Hull and his revisers is at all suitable. It should be noted that in Hull's own statements (1943) the relationship between experimental variables and intervening variables is usually anything but linear. If the exact form of the functional relationship is not known, performing linear algebraic operations on the intervening variables is practically meaningless. Under these conditions, for example, one cannot prove on the basis of experimental data whether changes in response strength are the result of an additive or a multiplicative relationship between intervening variables. From more fundamental considerations, Hilgard (1958) points out that Hull's intervening variables cannot in their present form be multiplied meaningfully, since they are not in comparable units of measurement. Certainly the least objectionable formula for reaction potential is also the least specific. Consequently it has the least predictive power:

$$\bar{E}_R = f(D, K, sH_R, I_R, \text{etc.})$$

In view of the facts here noted, great difficulties arise when Hull and his revisers become more explicit about

the nature of the relationships between these variables.

Though it would not be in keeping with the spirit of Hull's formal theorizing, some of the problems might be avoided if Hull's formula for  $s\bar{E}_R$  were regarded, not as a true mathematical equation, but merely as a kind of shorthand for expressing certain relationships suggested by empirical findings. The arithmetic signs of addition, subtraction, and multiplication in the formula would then not be taken too literally. Thus,  $E = H - I$  would not be taken to mean that inhibition subtracts from habit and that when  $E$  finally equals zero, the habit has been removed and the organism restored to the same state as before the habit had been acquired. The equation merely states in shorthand form that reaction potential, as inferred from some measure of response strength, decreases as the experimental procedures said to increase habit strength are removed and the conditions said to produce inhibition are applied. The subtraction sign is used here, not in a strict mathematical sense, but only as a shorthand expression for an experimental manipulation. Whether Hull has chosen to add or to multiply various intervening variables most likely has been a result of his attempt primarily to represent known empirical relationships rather than to maintain logical consistency within his theory. He most likely formulated  $D \times sH_R$ , for example, because he believed this interaction of habit and drive represented the experimental evidence. And most probably the reason he did not formulate  $D \times sI_R$ , even though his theory seems to call for this logically, was simply because he found no evidence that suggests an interaction between drive and inhibition.

From the foregoing considerations, probably the ultimate conclusion to which we are forced regarding the attempted revisions of Hull's theory is not so much that these revisions are no improvement over Hull, but that it is futile to attempt to improve upon Hull by mere juggling of his intervening variables. Hullian theory will not be improved by continuing to work with the concepts of drive, habit, inhibition, etc. in exactly the same form they were given by Hull. The very building blocks of the theory, so to speak, are inadequate, and no amount of recombining them in new ways is likely to result in any substantial advance in learning theory.

#### REFORMULATIONS AND EMPIRICAL EVIDENCE

##### $sH_R - sI_R$

While Hull (1943) refers to  $sI_R$  as a "negative habit," there is no indication in his writing that he regards  $sI_R$  as merely negative  $sH_R$ . The revisions suggested by Osgood and by Jones are based on the assumption that  $sH_R$  and  $sI_R$  are basically the same phenomenon,  $sI_R$  merely being the negative counterpart of  $sH_R$ . Thus, if they are the same process but merely opposite in effect, it seems logical that one should subtract from the other. Similarly, if  $sH_R$  interacts with drive, so should  $sI_R$ . Hull, however, quite clearly did not regard  $sH_R$  and  $sI_R$  as basically one and the same phenomenon, and his reasons are based on experimental evidence that reveals differences between the two. Pavlov (1927) originally pointed out the greater susceptibility of internal inhibition (of which  $sI_R$  is one variety) to external inhibition (i.e., disinhibition) than is the case with the excitatory process corresponding to Hull's  $sH_R$ . That  $sI_R$  is more labile

and sensitive to external influences than is  $sH_R$  suggests that it is not merely the negative counterpart of the same phenomenon. Therefore, Hull is consistent with Pavlov in not subtracting  $sI_R$  directly from  $sH_R$ .

Another line of evidence that excitation (conditioning) and inhibition (extinction) are basically different processes is well demonstrated in a series of experiments by Reynolds (1945a, 1945b), which showed that acquisition of a conditioned response is slower for massed than for distributed trials, while the *reverse* relationship holds for extinction. Also a number of studies (Hilgard & Marquis, 1940, p. 119) have shown a *negative* correlation between the speed of conditioning and of extinction.

The issue of whether the generalization gradients of excitation (conditioning) and inhibition (extinction) are the same or different was left undecided by Hull (1943, p. 265). The Bass and Hull (1934) and Hovland (1937) studies referred to by Hull were not adequate to answer this question. Not finding evidence to the contrary, Hull merely assumed that the generalization gradients of excitation and inhibition were the same, which is a convenient assumption in his theory of simple discrimination learning (1943, p. 267) based on the interaction of the gradients of excitation and inhibition. On this point, however, there is now some tentative evidence that seems to contradict Hull's assumption. Liberman (1951) found that extinction ( $sI_R$ )<sup>1</sup> has broader transfer

<sup>1</sup> In Hull's system, though the entire process of extinction is not explained in terms of only  $sI_R$ , but includes reactive inhibition ( $I_R$ ) as well, once extinction is complete, or after enough time (probably 5 to 10 minutes) has elapsed for the dissipation of  $I_R$ , extinction is conceived of as solely a function of the relative magnitudes of the positive reaction potential and  $sI_R$ .

effects than acquisition ( $sH_R$ ). Also there is some evidence (Razran, 1938) that the stimulus generalization of extinction ( $sI_R$ ) differs from that of excitation ( $sH_R$ ), in that extinction shows greater stimulus generalization; the gradient of its generalization contains fewer steps; the stimulus generalization of extinction, unlike that of acquisition, does not extend to heterogeneous CRs; and generalization of extinction is more affected by drugs than is generalization of conditioning.

The formulation  $sH_R - sI_R$  seems misleading in view of the fact that successive periods of acquisition and extinction become more rapid and that an organism in which an acquired response has been extinguished is not the same as an organism that had never acquired the response. Razran (1956) has pointed out that in a partially extinguished CR there can be shown the co-existence of two opposing processes, positive and negative. "Even a wholly extinguished CR bears, by all signs, within itself a two-way CR connection" (p. 42). Successive acquisition and extinction may be conceived of as a kind of discrimination learning, in which both  $sH_R$  and  $sI_R$  grow simultaneously, neither one diminishing the other. The cessation of reinforcement becomes a cue, a conditioned inhibitor, the strength of which increases throughout successive extinction periods (Bullock & Smith, 1953; Perkins & Cacioppo, 1950). This kind of discrimination learning is likely to be a very primitive kind of discrimination not involving symbolic or mediating processes. Tentative evidence for this opinion is found in the experiments on spinal conditioning, which, however, are not yet entirely beyond dispute as examples of true conditioning. Nevertheless, for what it is

worth, Shurrager and Shurrager (1946) have reported that both conditioning and extinction, measured at a single synapse in a spinal preparation, become faster with successive periods of conditioning and extinction.

Hull (1952, p. 114) also pointed out that the delay CR (the "inhibition of delay" being due to  $sI_R$ ) is eliminated by certain drugs, for example, caffeine and benzedrine. It is hard to see why the CR itself would not be markedly weakened or eliminated altogether if these drugs affected both  $sH_R$  and  $sI_R$  in the same manner. The CR is strengthened, however, while the period of delay is markedly shortened. Certain drugs thus seem to have opposite effects on  $sH_R$  and  $sI_R$ , suggesting again that they represent essentially different underlying physiological processes. Skinner's (1938, pp. 412-413) finding that benzedrine and caffeine increase the number of responses to a criterion of extinction lends plausibility to the idea that these drugs have different effects on  $sH_R$  and  $sI_R$ . If  $sH_R$  and  $sI_R$  were the same process, then a drug increasing  $sH_R$  would also increase the inhibitory effect of each nonreinforced response. If this were the case, the unfailing effect of stimulant drugs in increasing the number of responses to extinction could not easily be accounted for. The evidence bearing on this subject, however, is not crucial, in that we do not have evidence regarding the *percentage* increase in responding during extinction under benzedrine *over the operant level* (preconditioning response rate) under benzedrine. Also it should be noted that the theoretical problem hinges to some extent upon the hypothesized relationship between excitation (or  $sH_R$ ) and inhibition ( $sI_R$ ); that is, whether it is the absolute *difference* between the

two that matters or the *ratio* (or "balance") between excitation and inhibition. In the Pavlovian system it is the balance or ratio of excitation to inhibition that determines reaction potential. In Hull's system it is the absolute difference between  $sH_R$  (and the variables interacting with it) and  $I_R$ . A strictly Pavlovian revision of Hull might take the following form:

$$sE_R = \log \frac{D \times sH_R}{I_R}$$

Thus it is the balance between excitatory and inhibitory processes that is emphasized and not the absolute difference. In this equation, when the total inhibitory potential ( $I_R$ ) is equal in strength to  $D \times sH_R$ , the ratio of  $D \times sH_R/I_R$  becomes 1.0, and since  $\log 1.0 = 0$ , the effective reaction potential ( $sE_R$ ) will equal zero.

The fact that Eysenck and his co-workers have subscribed to the Jones revision would seem incompatible with Eysenck's (1956) theory concerning the extinction of  $sI_R$ . The extinction of  $sI_R$  is paradoxical and inconsistent with other aspects of Hull's theory and also of Jones' revision. If, as maintained by Jones and by Eysenck,  $sI_R$  is merely negative  $sH_R$ , then the mere lack of reinforcement of  $sI_R$  (reinforcement being the dissipation or avoidance of  $I_R$ ) should not result in a decrease in  $sI_R$ . Lack of reinforcement does not diminish the  $sH_R$  already present, so it should not diminish  $sI_R$  either. The notion that extinction is an active process of an increasing inhibition ( $I_R$ ) depressing performance ( $sE_R$ ) is basic in Hull's system. It, therefore, seems absurd, while remaining in the Hullian framework, to speak of the extinction of inhibition without first postulating a sec-

ond inhibitory process which depresses the first. Fortunately, there is no experimental evidence at present to suggest that such a complication would be necessary.

#### $D \times sI_R$

In Hull's theory there is no interaction between drive and conditioned inhibition. The  $D \times sI_R$  interaction, however, is explicit in a number of the revisions. Since  $sI_R$  is the primary and essential intervening variable accounting for experimental extinction, we may well examine the different predictions generated by Hull and the revisions with respect to the  $D \times sI_R$  interaction.

According to Hull, since  $D$  multiplies only  $sH_R$  and not  $sI_R$ , we should predict that certain measures of extinction will be affected by changes in  $D$ . With the Hullian formula  $D \times sH_R - sI_R$ , one can predict that under a high drive level there will be a greater number of responses to extinction ( $n$ ) than under low drive. The same increment of  $sI_R$  is generated by each response during extinction, regardless of the level of  $D$ , while the positive reaction potential ( $D \times sH_R$ ) is increased by a higher level of  $D$ . Not only does it follow from Hull's formula that a greater number of responses is required for extinction, but extinction curves under high and low  $D$  should be parallel. They approach the criterion of extinction with the same slope, but reach it at different points.

The revisions containing the  $D \times sI_R$  interaction generate predictions that are exactly opposite to the foregoing. If net reaction potential is a resultant of  $D \times sH_R - D \times sI_R$ , then every increment of  $sI_R$  will be increased by  $D$  to the same degree that  $sH_R$  has been increased. Consequently, there should be the

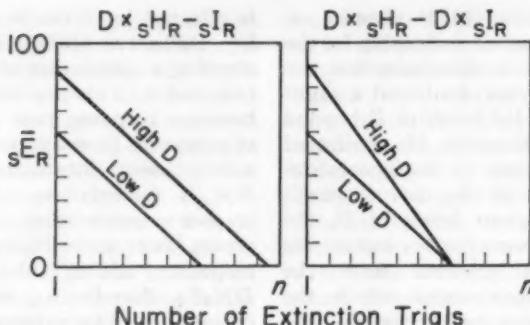


FIG. 1. The relationships between drive ( $D$ ), number of trials to extinction ( $n$ ), and effective reaction potential ( $sE_R$ ) as predicted by Hull's formulation (left) and by Jones's formulation (right).

same number of responses to extinction under high drive as under low drive. Also, the slopes of the extinction curves, as measured by, say, rate of responding, would be different under high and low drive. In other words, the curves would approach the criterion of extinction with *different slopes*, but would reach it at the *same point*.

If the proponents of the  $D \times S_I_R$  formulation object to the foregoing predictions on the grounds that  $I_R$  has not been taken into account, let it be pointed out that  $S_I_R$  is essential for complete extinction of the response and that extinction can take place with sufficiently spaced trials to prevent the growth of  $I_R$ . If, as Hull hypothesized (1943, pp. 300-301), the formation of  $S_I_R$  is dependent upon nonresponding being coincident with the dissipation of  $I_R$ , extinction could not take place if all  $I_R$  had dissipated in the interval between each presentation of the non-reinforced CS. Yet extinction is known to occur even with long inter-trial intervals of 24 hours or more, when  $I_R$  should supposedly have been completely dissipated (Razran, 1956, p. 43). This, along with the

fact that in all of the revisions an increment of  $I_R$  will reduce  $sE_R$  by the same proportion regardless of the level of  $D$ , makes  $I_R$  irrelevant to the present argument. (The  $D - I_R$  formulation is discussed at a later point.)

There is a considerable amount of experimental evidence bearing on the above predictions. The preponderance of evidence favors the Hullian formula and fails to support the notion of a  $D \times S_I_R$  interaction. Perin (1942), working with rats, found a marked positive relationship between  $D$  (degree of hunger) at the time of extinction and the number of responses required for extinction. Brandauer (1953) extinguished bar pressing in rats under three levels of drive (thirst) and found a positive relationship between strength of drive and number of responses during extinction. Even under minimal differences in hunger drive (.5, 1, 2 hours' deprivation) Saltzman and Koch (1948) found highly significant differences in number of responses to extinction in a modified Skinner box. Brown (1956) also found that rats on high drive make more responses during extinction than those on low

drive. Cautela (1956) showed essentially the same relationship for the extinction of a discrimination response. However, he found a slight decrease in  $n$  for levels of  $D$  beyond 23 hours' deprivation. He attributed this phenomenon to the generalization gradient of the drive stimuli; under the highest levels of  $D$ , the drive stimuli were further out on the generalization gradient from the drive conditions under which the original learning had occurred. The energizing and stimulus properties of drive are thus apt to interact in this type of experiment.

In experiments with human subjects, where anxiety has been used as a measure of drive, a similar relationship with extinction has been found. In one study, high anxiety subjects required almost twice the number of trials to extinguish the conditioned eyeblink as did low anxiety subjects (Spence & Farber, 1953). Bitterman and Holtzman (1952) obtained similar results in extinguishing the PGR in high and low anxiety subjects.

Skinner's (1938) early notion of the "reflex reserve" appears to be consistent with the  $D \times sI_R$  formulation. Skinner believed that the number of responses emitted during extinction was solely a function of the number of previously reinforced responses and the schedule of reinforcement. Thus drive should not affect  $n$ , but would affect only the rate of emission of responses. The reflex reserve concept, however, has long since been found unfruitful. While theoretically it is probably not a strictly testable hypothesis, it now at least appears quite incorrect in view of the evidence (Ellson, 1939). Skinner's (1938) original belief that rate of responding, but not the number of responses in extinction ( $n$ ),

is affected by drive is contradicted by Bullock's (1950) investigation showing a correlation of .61 between rate and  $n$ . This positive correlation between response rate and number of responses to extinction would certainly seem inconsistent with a  $D \times sI_R$  formulation. If drive increases response rate,  $sI_R$  should increase faster under higher drive, each response adding the increment  $D \times sI_R$ , thus leading to more rapid extinction. The evidence is exactly the contrary. Higher drive not only increases the rate of response, but also increases the total number of responses to a criterion of extinction.

The best available evidence also indicates that the slope of the extinction curve is the same under high and low drive, as would be predicted from Hull's theory. Sackett (1939) showed that when the extinction curves of two groups of rats, one group extinguished under 6 hours' hunger drive and the other under 30 hours' drive, are Vincentized, the forms of the two curves are almost identical. The 30-hour group produced more responses to extinction and required more time to extinguish, but the slope of the extinction curve was the same as that of the 6-hour group. Barry (1958) trained rats in a running response and extinguished them under high and low drive. The extinction curves were parallel, and when drive was equalized in both groups late in extinction, the curves converged and were identical after three trials. When drive was equal for both groups early in extinction, and then, later in extinction, the groups were run under high and low drive, the extinction curves diverged, and, after three trials, continued almost parallel, as would be predicted from Hull. (The fact that it took three trials, rather than one,

for the curves to converge or diverge after the change in  $D$ , however, is somewhat embarrassing to Hull's theory as it is also to the revision.) Both these findings are consistent with the  $D \times sH_R - sI_R$  formulation and not with  $D \times sH_R - D \times sI_R$ . But these experiments cannot be regarded as at all definitive in view of the finding of Reynolds, Marx, and Henderson (1952) of an interaction between  $D$  and the incentive factor  $K$  (a function of amount of reward). This interaction plays havoc with any theoretical conclusions drawn from experiments on the effects of drive on extinction in which the incentive factor has not been taken into account. Reynolds et al. (1952) had four groups of rats learn bar pressing under all combinations of high drive-low drive and large reward-small reward. All animals were given extinction trials under equal drive. It was found that

in those learning situations where a relatively large amount of reward is employed for reinforcement, high  $D$  animals extinguish more readily than low  $D$  animals; and . . . where a relatively small reward is given per reinforcement, low  $D$  animals extinguish more readily than high  $D$  animals (pp. 41-42).

Hull's theory and its revisions generate conflicting predictions regarding spontaneous recovery. In the Jones (1958) formula,  $s\bar{E}_R = D - I_R$   $\times (sH_R - sI_R)$ , spontaneous recovery could occur only if at the end of the first set of extinction trials  $D - I_R = 0$ . But this formulation would lead to problems, since, if  $D - I_R = 0$ , no habits at all could be activated temporarily until some of the  $I_R$  had dissipated, and no behavior of any kind would occur after the end of the first extinction period. We know very well, however, that animals go on behaving in various ways immediately following the extinction of

a particular response. But then if we do not wish to assume that  $D - I_R$  is equal to zero immediately after the first extinction period, we must assume that  $sH_R - sI_R$  equals zero, or extinction would not have occurred. Yet if  $sH_R - sI_R$  were zero, there could be no spontaneous recovery. Conceivably one way out of this dilemma for the Jones revision is to make some assumptions about a reaction threshold which must be exceeded before an overt response is made. Thus, overt extinction could occur before either  $D - I_R = 0$  or  $sH_R - sI_R = 0$ . Spontaneous recovery would then result from the dissipation of  $I_R$ , as in Hull's theory. If this were true, one might predict from the Jones revision that there would be very little, if any, spontaneous recovery after extinction under high drive, but greater amounts of spontaneous recovery after extinction under low drive. Since  $D - I_R$  would approach the threshold value quickly where  $D$  is initially low, there would result an appreciable increase in  $D$ , and hence of response strength, with the dissipation of  $I_R$ , and spontaneous recovery would result. Under high drive  $D - I_R$  would not approach the threshold value as quickly as would  $sH_R - sI_R$ . Thus, since  $sH_R - sI_R$  would be a smaller value after the first extinction, there should be less spontaneous recovery at the beginning of subsequent extinction periods.

Different predictions may be made from Hull and the  $D \times sI_R$  revision concerning the effect of an increase in drive after extinction is complete. According to Hull's  $(D \times sH_R) - sI_R$ , an increase in drive after complete extinction should result in further "spontaneous recovery." According to the  $D \times (sH_R - sI_R)$  formulation, once extinction is complete (i.e.,

$sH_R - sI_R = 0$ ), an increase in  $D$  should not produce any "spontaneous recovery."

Unfortunately, the experimental evidence bearing on all these predictions is meagre, conflicting, and inconclusive. Hull (1943, p. 249) cites Pavlov's finding that an increase in drive after extinction is complete causes the reappearance of the CR in the presence of the CS. This is, of course, consistent with Hull's formulation, but not with the  $D \times sI_R$  formulation. The same phenomenon seems to occur also in instrumental conditioning. Jenkins and Daugherty (1951) extinguished a pecking response in pigeons under three levels of drive. They found that the number of responses in extinction is a function of drive level and that when extinction was relatively complete an increase in drive caused gross recovery of the conditioned behavior. The authors used the term "relatively complete" extinction because the pecking response in pigeons never seems to be completely extinguished. But the recovery of a "relatively extinguished" CR under increased drive is certainly more consistent with  $(D \times sH_R) - sI_R$  than with  $D \times (sH_R - sI_R)$ . The writer knows of only one study that appears to contradict the finding of Jenkins and Daugherty. Crocetti (1952) found that when rats were "completely" extinguished in a Skinner box, increase in drive did not increase the response rate over the pre-conditioning response rate under the higher level of drive. (Extinction was considered complete when the response rate became equal to the operant level prior to conditioning.) This finding is, of course, inconsistent with Hull's  $(D \times sH_R) - sI_R$ . Crocetti did not control for the changes in the drive stimulus ( $S_D$ ) with in-

creased hunger, and so his finding is not definitive with respect to the present theoretical issue. If we assume that  $sH_R$  and  $sI_R$  are conditioned to  $S_D$  as well as to other stimuli, then the changes in  $S_D$  from the conditioning trials to the extinction trials or spontaneous recovery trials becomes a crucial point in this type of experiment. Fortunately in an experiment by Lewis and Cotton (1957) the effect of such changes in  $S_D$  was taken into account. Three groups of rats were trained in a running response under three levels of drive, viz., 1, 6, and 22 hours' food deprivation. Each group was then divided into three groups which underwent extinction under 1, 6, and 22 hours' drive. Extinction proceeded more rapidly under lower drive, as would be expected from Hull's formulation, but drive level seemed to have no effect on the magnitude of spontaneous recovery, a fact which is inconsistent with  $(D \times sH_R) - sI_R$ . But the  $D \times (sH_R - sI_R)$  revision cannot comprehend both of these findings either, for with this formulation drive level should have no effect on number of trials to a criterion of extinction. It seems obvious that clarification of the effects of drive on spontaneous recovery must await further experimentation which is specifically designed for this purpose and which takes into account both the energizing and stimulus properties of drive. Some of the lack of consistency and agreement in this area may also be due to interspecies differences and to the use of different measures of response strength. Latency, running time, response rate, and number of trials to extinction are used singly in different studies as measures of response strength even though they are far from being perfectly correlated. Each measure un-

doubtedly involves certain parameters peculiar to itself. To use only one such measure of response strength and only one species of animal is an inadequate method for testing a precise deduction from a general behavior theory.

In the delayed CR, the inhibition of the response during the period of delay is attributed in the Hullian system to  $sI_R$  (Hull, 1952, p. 114). Consistent with Hull's formulation of  $D \times sH_R - sI_R$  is the fact that an increase in  $D$  lessens or eliminates the period of delay in the CR. The  $D \times sI_R$  formulation does not accommodate this fact, but leads to an opposite prediction, i.e., an increase in  $D$  should strengthen the inhibition of delay.

One of the weakest points in Hull's system involves the dependence of  $sI_R$  upon  $I_R$ . It is no less troublesome to any of the revisions. (Spence excepted, since his inhibition concept has nothing in common with  $I_R$ .) It is stated that  $I_R$  is generated whenever a response is made, the amount of  $I_R$  being a function of the effortfulness of the response, and that  $I_R$  rapidly dissipates, accumulating only if responses follow one another in rapid succession. The dissipation of  $I_R$ , a "negative drive state," reinforces the habit of not responding, which is  $sI_R$ . This hypothesis encounters obvious difficulties. If a response is followed by the dissipation of  $I_R$ , this would seem to have all the requirements for reinforcing the response, leading to increased response strength rather than extinction.<sup>2</sup>

<sup>2</sup> One can get around this problem, of course, by invoking the gradient of reinforcement. If the time required for the dissipation of  $I_R$  is greater than the effective gradient of reinforcement, the foregoing proposition would not hold true. At present there is no basis for arguing the point. While Hull gives 20-30 seconds as the maximum delay between the

Also, subzero extinction would be unlikely if increases in  $sI_R$  were dependent upon *reactive* inhibition ( $I_R$ ). And it is almost impossible to explain the extinction of relatively effortless CRs, such as salivation, eyeblink, and the alpha rhythm, when the extinction trials are widely spaced. Pavlov (1927, p. 76) obtained rapid extinction of the salivary CR using only one presentation of the CS per day. Razran (1956, p. 43) has reviewed the evidence that contradicts a theory of extinction based on reactive inhibition. There are even cases where spaced trials have led to more rapid extinction than massed trials (Sheffield, 1950; Stanley, 1952). Kimble (1950) has argued from studies of motor learning that a certain threshold or critical level of  $I_R$  must be reached before  $sI_R$  develops. Motor learning experiments have presumably shown that  $I_R$  can form without leaving behind any  $sI_R$ . This is inconsistent with extinction based on widely spaced trials. In fact, it does not seem to the writer that the Hullian inhibition postulates, as they have been used in the field of motor learning, represent the same processes found in extinction phenomena. It has been a case of giving the same theoretical labels to basically different processes. The most fundamental difference between  $sI_R$  in conditioning and in motor learning has to do with the amount of response necessary to produce  $sI_R$ . Five or six minutes of pursuit rotor practice seems necessary before  $sI_R$  is in evi-

response and reinforcement if reinforcement is to be effective, the time required for the dissipation of  $I_R$  is solely a function of the amount of  $I_R$  generated by the response and, therefore, is variable, although the rate of dissipation of  $I_R$  may not be variable. Perhaps an even simpler way out is the idea that  $I_R$  leads to a "resting response" which in turn is reinforced by the dissipation of  $I_R$ .

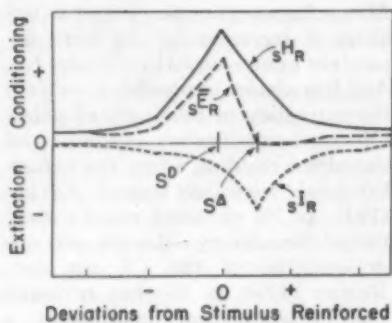


FIG. 2. Illustrates algebraic summation theory of discrimination. (Effective reaction potential,  $sE_R$ , is a result of subtraction of generalized extinction,  $sI_R$ , from generalized conditioning,  $sH_R$ . See text for full explanation.)

dence, while only a single conditioned response, such as salivation, PGR, or eyeblink, is evidently sufficient to produce  $sI_R$ . Thus it does not seem that the  $sI_R$  invoked in theories of motor learning could be the same  $sI_R$  as that in Hull's theory of conditioning.

It is also held by Hull, and even more explicitly by his revisers, that the amount of  $sI_R$  built up per trial is related to the amount of  $I_R$  dissipated, the dissipation acting as reinforcement for the negative habit,  $sI_R$ . But this is inconsistent with Hull's own revision of his theory (Hull, 1951), in which the growth of habit,  $sH_R$ , and presumably also of negative habit,  $sI_R$ , is a function only of the *number* of reinforcements and not the *amount* of reinforcement. None of these awkward predicaments has been remedied by the revisions here reviewed. Those revisions insisting on the theoretical equivalence of  $sH_R$  and  $sI_R$  as being merely positive and negative habits have retained one of the weakest elements in the Hullian system.

*Discrimination learning.* If dis-

crimination learning involves an increase in habit strength to the positive stimulus ( $S^D$ ) and an increase in inhibition ( $I_R$  and  $sI_R$ ) to the negative stimulus ( $S^A$ ), then the effects of drive on discrimination learning should be highly germane to the plausibility of the  $D \times sI_R$  formulation. Jones (1958) invokes Spence's (1937) theory of discrimination learning, adapted by Hull, involving the overlapping generalization gradients of  $sH_R$  and  $sI_R$ , in support of the  $D \times sI_R$  part of his revision. This theory is illustrated in Figure 2. The discrimination would be perfect (except for behavioral oscillation) when the net reaction potential resulting from subtracting the generalized  $sI_R$  from the generalized  $sH_R$  is some positive value for  $S^D$  and zero for  $S^A$ , as in Figure 2. Jones (1958) argues that, according to Hull's  $D \times sH_R - sI_R$ , an increase in  $D$  should obliterate the learned discrimination. Since some discriminations are not obliterated or even weakened by an increase in  $D$ , Jones reasons that  $sI_R$  must also be multiplied by  $D$ , so that the increase in  $sI_R$  will be proportional to the increase of  $sH_R$  when multiplied by  $D$ , thereby preserving the discrimination.

Before Jones' argument can be evaluated, some clarification of the Spence-Hull theory of discrimination learning is necessary. In the first place, there is often confusion concerning whether discrimination learning is a matter only of the *relative* strengths of  $sE_R$  to the  $S^D$  and  $S^A$ , or whether the formation of a discrimination requires the reduction of  $sE_R$  to the  $S^A$  to zero or at least below the operant level of the response, i.e., below the strength of the response before any conditioning or extinction has occurred. If the former,

then all that would be necessary for discrimination to take place would be that the  $S^D$  have greater  $s\bar{E}_R$  than the  $S^A$ . The  $s\bar{E}_R$  to the  $S^A$  would not necessarily have to undergo some degree of extinction. If this were the case, Jones' use of the Spence-Hull theory of discrimination, as illustrated in Figure 2, would not be applicable to the present argument concerning the effects of drive on discrimination learning. The evidence, however, strongly suggests that the  $s\bar{E}_R$  to the  $S^A$  must undergo some degree of extinction for discrimination to become nearly perfect. To this extent, at least, the Spence-Hull theory appears to be correct.

For example, Grice (1948) gave one group of rats 200 rewarded trials in responding to a disc 8 centimeters in diameter and gave another group of rats 200 rewarded trials in responding to a 5-centimeter disc. Then both groups were given discrimination training, with the 8-centimeter disc as the  $S^D$  and the 5-centimeter disc as the  $S^A$ . The group which had been previously rewarded on the 8-centimeter disc learned the discrimination faster. Now if all that were involved in discrimination were the *relative* response strengths to the  $S^D$  and  $S^A$ , the 8-centimeter group should have learned to make the discrimination immediately, since response to the  $S^D$  had already been rewarded on 200 trials, and the response strength to the  $S^A$  resulting from stimulus generalization would have been less than the response strength to the  $S^D$ . Since the learning curve for the acquisition of the discrimination is very gradual, however, it suggests that *extinction* of the response to the  $S^A$  through non-reinforcement is necessary for the learning of the discrimination.

An even more cogent demonstra-

tion of the necessity for extinction of  $S^A$  in discrimination learning is an experiment by Fitzwater (1952). Three groups of rats were used: Groups A, B, and C. In preliminary training Group A was run an equal number of times into each of two alleys having differential cues—call them  $X$  and  $Y$ , respectively.  $X$  was always reinforced;  $Y$  was never reinforced. Group B was run an equal number of times into each of two alleys having the Cues  $X$  and  $Z$ .  $X$  was always reinforced;  $Z$  was never reinforced. Group C was run only into one alley, with Cue  $X$ , the same number of times as the other groups. Then discrimination training began, with the animals having to learn to discriminate  $X$  as the  $S^D$  and  $Y$  as the  $S^A$ . Group A learned the discrimination most rapidly, while Groups B and C did not differ significantly in speed of learning. The theoretical interpretation of these results is that Group A had already built up inhibition to the  $S^A$ , while Groups B and C had not. Fitzwater concluded that

apparently in a visual discrimination task it is about as important to establish an avoidance habit as an approach habit, and that an appreciable discrimination does not seem to occur if an approach habit is established alone (p. 480).

The terms "approach habit" and "avoidance habit" may be interpreted in the context of the present discussion as excitation (or  $sH_R$ ) and extinction ( $sI_R$ ), respectively. Thus it is apparent that a decrease in  $s\bar{E}_R$  to the  $S^A$  as well as an increase in  $s\bar{E}_R$  to the  $S^D$  is necessary for discrimination learning. It is not just a matter of  $s\bar{E}_R$  to the  $S^D$  being relatively greater than to the  $S^A$ .

Another experiment by Grice (1949) offers further evidence that discrimination depends upon the *extinction* of the response to the  $S^A$  and

not merely a relative difference in response strengths between  $S^D$  and  $S^A$ . One group of rats was trained in a visual size discrimination with  $S^D$  and  $S^A$  presented simultaneously, and another group was trained on the same discrimination with  $S^D$  and  $S^A$  presented successively in random order. Grice found no difference between the "simultaneous" and "successive" groups in the rate of learning the discrimination. In both cases learning apparently consisted of increasing the response strength to the  $S^D$  and completely extinguishing the response to  $S^A$ . Furthermore it was found that the rats which learned the problem as a pair (i.e., simultaneous presentation) responded differently to the  $S^D$  and  $S^A$  when they appeared singly, showing that even under simultaneous presentation of the  $S^D$  and  $S^A$ , the response to the  $S^A$  had undergone extinction.

It is not maintained that *complete* extinction of the response to  $S^A$  is necessary. Extinction is a relative matter and is probably best measured, not in relation to some theoretical "absolute zero," but in relation to the "operant level" or probability of occurrence of the particular response before extinction trials have been assumed to take place. In the Grice (1949) experiment there was a decrease in latency of response to  $S^D$  and an increase in latency of response to  $S^A$ , whether the two stimuli were presented simultaneously or successively.  $s\bar{E}_R$  to the  $S^A$ , as measured by latency, was considerably less at the end of discrimination training than at the beginning. In fact, extinction of response to  $S^A$  may play a greater role in discrimination learning than does the strengthening of the response to  $S^D$ . Webb (1950) trained rats to jump to a black-white discrimination until it

was well learned. When, after training, only the  $S^D$  was presented to the rats, the mean latency of their response was 2.0 seconds, which was not significantly less than the pre-training latency. On the other hand, when only the  $S^A$  was presented, the mean latency of response was 80.5 seconds, which may be interpreted as indicating considerable extinction or inhibition of the response to the  $S^A$ . If one defines the zero level of  $s\bar{E}_R$  in the Hull-Spence model in Figure 2 simply as the operant level (i.e., the pretraining latency or probability of responding to the particular stimuli), then this model appears to be quite consistent with the experimental evidence in showing that discrimination depends upon extinction of the response to the  $S^A$ .

This model, however, seems to be deficient in some other respects. Hanson (1957), for example, performed a very careful experiment which led to the conclusion that over-all response strength is *not* weakened by discrimination training, as would be predicted from the Spence-Hull model. (That is, since the resultant  $s\bar{E}_R$  is the algebraic sum of generalized excitation and inhibition,  $s\bar{E}_R$  to the  $S^D$  should be less after discrimination training than it would be in simple conditioning to a single stimulus.) Hanson concluded that

the major result of discrimination training is to bring a large proportion of the responses available in extinction under the control of another range of stimuli, those which do not ordinarily gain control of the response as the result of simple conditioning without differential reinforcement (p. 889).

This conclusion is not compatible with the Spence-Hull theory.

It may be argued that Jones has taken the Spence-Hull diagram (Figure 2) too literally. Very little is

known about the actual shapes of the generalization gradients of  $sH_R$  and  $sI_R$ , and until a proper metric is worked out, arguments over this point cannot be settled. What little evidence there is, though far from conclusive, suggests that the generalization gradients of excitation and inhibition are probably different in a number of respects (Razran, 1938). Furthermore, the amount of overlap of the gradients of excitation and inhibition will depend on the distance apart of  $S^D$  and  $S^A$ , and there is reason to believe that the effects of drive on discrimination will interact with the degree of disparity between  $S^D$  and  $S^A$  (Broadhurst, 1957). We would predict from Hull's  $D \times sH_R - sI_R$  that the farther apart  $S^D$  and  $S^A$  are, the less deleterious to the discrimination are the effects of increased drive. This essentially is the Yerkes-Dodson Law (Yerkes & Dodson, 1908), which, in its most general form, states that the optimum motivation for a learning task decreases with increasing difficulty. This relationship between drive and difficulty of discrimination, however, cannot be predicted from the Jones formulation of  $D \times (sH_R - sI_R)$ .

Rather than arguing from a highly hypothetical model involving the relative shapes and magnitudes of the generalization gradients of  $sH_R$  and  $sI_R$ , as Jones has done, we can better make predictions concerning the directly observable effects of increased drive on discriminations. What is the effect of drive on the initial learning of a discrimination, and does an increase in drive have a different effect on the learning of easy and difficult discriminations, as determined by time required for learning? What is the effect of change in drive on discriminations that are already established? What effect does

a change in drive have on the extinction of a discrimination?

In discrimination learning, since the relative amounts of  $sH_R$  and  $sI_R$  built up to the  $S^D$  and  $S^A$  are different, we would expect from the  $D \times (sH_R - sI_R)$  formula that an increase in  $D$  would always have a facilitative effect on learning a discrimination. The degree of facilitation would depend upon the degree of difference between  $S^D$  and  $S^A$ . If we assumed considerable overlapping of generalization gradients, then there would be relatively little effect of an increase in  $D$ . If the discrimination were easy, increases in  $D$  should improve the discrimination, since the relatively greater  $sH_R$  to the  $S^D$  and the relatively greater  $sI_R$  to the  $S^A$  would both be multiplied by  $D$ . In no case should discrimination be weakened by an increase in  $D$ .

On the other hand, if we assume that response to  $S^A$  must undergo extinction for a discrimination to be learned, Hull's formula  $D \times sH_R - sI_R$  leads to quite different predictions, viz., that increase in  $D$  should weaken difficult discriminations, where one might assume overlap of the stimulus generalization gradients, but would strengthen discriminations in which  $S^D$  and  $S^A$  are widely separated on the generalization gradient.

What is the evidence? We have already mentioned the Yerkes-Dodson Law, which is possibly consistent with Hull, but certainly not with the  $D \times (sH_R - sI_R)$  formula. Broadhurst (1957) has demonstrated this "law" most effectively, using rats in a brightness discrimination problem and manipulating drive by means of oxygen deprivation. Skinner (1938, p. 188) has observed that it is important in establishing discriminant operant conditioning to keep the hunger drive as constant as possible,

for changes in drive disturb the discrimination. More explicitly, Teel (1952) has shown that in selective learning, in which correct responses are reinforced and incorrect responses are nonreinforced or extinguished, rats under high drive (food deprivation) require a *greater* number of trials to reach a criterion of learning than rats under low drive. One cannot predict these facts from the  $D \times (sH_R - sI_R)$  formula. In fact, just the opposite outcome would be predicted for the Teel experiment. With human subjects, Hilgard, Jones, and Kaplan (1951) found that high anxiety subjects (on Taylor Manifest Anxiety scale) had greater difficulty than low anxiety subjects in forming a *discriminatory CR*. It is well-established that anxious subjects develop simple eyeblink CRs more readily than nonanxious subjects. (This relationship has not been found to hold, however, for autonomic CRs.) Interpreting anxiety as a drive, both sets of findings are consistent with Hull, but not with  $D \times (sH_R - sI_R)$ . An experiment by Spence and Farber (1954) found that the difference between high and low anxious subjects in forming a discriminatory response showed up only on the  $S^D$  but not on the  $S^A$ . That is,  $D$  (anxiety) seemed to affect only the CS (i.e.,  $S^D$ ) associated with relatively greater  $sH_R$  and not the CS (i.e.,  $S^A$ ) associated with relatively greater  $sI_R$ . Spence interprets this finding as evidence that  $D$  interacts with excitation ( $sH_R$ ) but not with inhibition ( $sI_R$ ).

In a well-established discrimination, in which  $S^D$  and  $S^A$  are relatively far apart on the stimulus generalization gradient, and in which relatively more  $sH_R$  than  $sI_R$  has been built up to  $S^D$  than to  $S^A$ , and relatively more  $sI_R$  built up to  $S^A$

than to  $S^D$ , we would predict from  $D \times (sH_R - sI_R)$  an improvement in the discrimination with an increase in drive. That is, the ratio of number of responses to  $S^D$  to number of responses to  $S^A$  should increase, since response to  $S^D$  is increased by  $D \times sH_R$ , and inhibition of response to  $S^A$  is increased by  $D \times sI_R$ . Dinsmoor (1952) performed an experiment bearing on this point. A simple discrimination habit was well-established in rats in the Skinner box, with  $S^D$  being the presence of light and  $S^A$  being total darkness. When  $D$  was increased to varying degrees by food deprivation, the *number* of responses per unit of time to both  $S^D$  and  $S^A$  increased, but the *ratio* of  $S^D$  and  $S^A$  responses remained exactly the same at seven different degrees of hunger. In short, the discrimination was not improved by an increase in  $D$ . Though Hull's theory is not sufficiently quantified to have precisely predicted the outcome of this particular experiment (because absolute levels of  $D$  and  $sH_R$  as well as the jnd's between  $S^D$  and  $S^A$  must be taken into account), at least the result is consistent with the  $(D \times sH_R - sI_R)$  formulation.

There is other experimental evidence, however, which suggests that both the Hullian and the revised formulations are inadequate to explain the effects of drive on discrimination learning. A number of studies have found no relationship at all between drive and proficiency in selective learning or solving discrimination problems (Meyer, 1951; Miles, 1959; and a number of doctoral dissertations reported by Spence, Goodrich, & Ross, 1959). Spence et al. (1959) have scrutinized the conflicting findings in this field with a view to discovering the reason for the lack of agreement between various investiga-

tions on the effect of drive on selective learning and discrimination. They arrived at the hypothesis that performance in selective learning (such as learning a black-white discrimination) is independent of drive level when responses to the  $S^D$  and  $S^A$  are equated, but varies with drive when responses are not equated. They performed a set of experiments which supported this hypothesis. The results are inexplicable in terms of Hull's theory or any of its revisions except that of Spence. These findings suggest that the growth of  $sH_R$  is not a function of number of reinforced responses, as in Hull's system, but is a function merely of the number of responses, whether reinforced or not. The growth of inhibition is a function only of the number of nonreinforced trials. This formulation will account for the major finding of the experiment by Spence et al. (1959). But another aspect of their findings remains inexplicable in terms of any current theory of learning. When responses to  $S^D$  and  $S^A$  were equated, an increase in drive increased the response strength to *both* the  $S^D$  and  $S^A$ . But when the rats were forced to respond twice as often to  $S^D$  as to  $S^A$ , an increase in drive *increased* the response strength to  $S^D$  but *decreased* response strength to  $S^A$ . Spence et al. concluded that

the results of the two (experiments) are in fundamental disagreement so far as the effects of drive differences on the strength of nonreinforced responses are concerned. It is perhaps obvious that we need to obtain much more knowledge than we now possess concerning the variables affecting the development of response decrement with nonreinforcement. Unfortunately, this problem has been badly neglected in conditioning experiments with the consequence that such an empirically based theory as the present one [i.e., Spence's theory] is weakest in this area (p. 15).

Though the present state of our

knowledge in this area does not permit any definite conclusion regarding the effects of drive on discrimination learning, it appears that no current theory is able to comprehend all the relevant facts now available.

But now let us ask: What happens when a discrimination is *extinguished* under various levels of drive? Cautela (1956) trained rats in a discrimination under 23 hours' food deprivation and then extinguished the discriminative response under 0, 6, 12, 23, 47, and 71 hours' deprivation. The criterion of extinction was failure to respond to either  $S^D$  or  $S^A$  within 3 minutes. Many more responses were required for extinction under high drive levels (23, 47, or 71 hours' deprivation) than under low drive (0, 6, or 12 hours). This result can be predicted from  $D \times sH_R - sI_R$ . On the other hand, it is difficult to see why a change in drive should have any effect on the number of responses to extinction if  $sH_R$  and  $sI_R$  are *both* increased or decreased proportionately by changes in  $D$ , as stated in the revised formula.

#### $D - I_R$

Since Hull referred to reactive inhibition ( $I_R$ ) as a "negative drive," he has been accused of logical inconsistency for adding a drive to a habit (i.e.,  $I_R + sI_R$ ) and the suggested remedy has been the obvious one, viz., to subtract  $I_R$  from  $D$ . But predictions from this formulation lead to empirical embarrassment. For example, when extinction is carried out under massed trials, and, after a period of rest, there is some spontaneous recover, we must assume, according to the  $s\bar{E}_R = (D - I_R) \times (sH_R - sI_R)$  formulation, that  $D - I_R = 0$  at the end of the first extinction period. For there would be no spontaneous recovery if it were  $sH_R - sI_R$

that had become equal to zero. Yet, according to Hullian theory (including the revisions), no behavior can occur unless  $D$  is greater than zero. And it is known that an animal at the end of extinction is far from being inactive. Only the extinguished CR becomes inactive, while other behavior in the animal's repertoire is immediately evident. Theoretically this could not be so if the drive component in the equation for reaction potential were zero.

Experimental evidence contradicting  $D - I_R$  is presented by Hull (1952, p. 50). A rat is trained to press either of two bars in different locations in a Skinner box to obtain food. During extinction the rat alternates its response from one bar to the other.  $I_R$  does not have to dissipate before the alternate bar can be pressed. This strongly suggests that  $I_R$  must be associated with the particular response, rather than cause a diminution in the total drive state, which in the Hullian system is an amalgam of all the organic needs of the moment and their associated "drive stimuli" ( $S_D$ ).

In an experiment highly relevant to this point, Smith and Hay (1954) took advantage of the great sensitivity to changes in drive of rate of responding in the Skinner box. As soon as operant conditioning had led to a stable response rate, a discriminatory stimulus was introduced, the  $S^D$  always being reinforced, the  $S^A$  never. During the learning of the discrimination, the number of responses to  $S^D$  increased while the number of responses to  $S^A$  decreased, but the *rate of responding remained constant*. If the extinction of  $S^A$  had involved  $D - I_R$ , there should have been the decrease in over-all response rate which is associated with lowered drive. On the other hand, this finding

is entirely consistent with Hull's formulation.

#### $I_R \times sI_R$

Here we have a formulation which, if the rules of algebra are followed religiously in manipulating Hullian variables, leads to a paradox—a positive addition to reaction potential resulting from the interaction of two inhibitory variables. Jones (1958) even goes on to say that the paradoxical outcome of  $I_R \times sI_R$  increasing  $sE_R$  might explain the "ultraparadoxical effect" described by Pavlov (1927). This might be called explanation by clang association.<sup>3</sup> It is difficult for the writer to understand why Jones and other revisers have so gratuitously regarded the minus sign as being permanently attached to  $I_R$  and  $sI_R$ . Though these quantities are subtracted from positive reaction potential, the negative sign is not necessarily an inherent part of these inhibition variables. Even if  $I_R$  and  $sI_R$  were multiplicatively related, there is no reason why their product could not be subtracted from the positive reaction potential.

The empirical evidence regarding the  $I_R \times sI_R$  interaction is far from satisfactory, for there is always an "out" via the possible interaction of all the other intervening variables in the system. But in terms of sheer plausibility—and that is all one can

<sup>3</sup> The "paradoxical" and "ultraparadoxical" effects observed by Pavlov, in which a weaker intensity of the CS will elicit a CR that had been extinguished to a stronger intensity of the CS, are probably best explained in terms of a generalization gradient on the stimulus intensity dimension. Because of the gradient, extintive inhibition built up to a CS of one intensity will not be sufficient to inhibit the CR to a CS of a different intensity, even though it be weaker. Or the effect may be explained as disinhibition caused by a "novel" stimulus—novel because the intensity is weaker than that of the original CS.

go on at present—it must be said that  $I_R \times sI_R$  is a weak formulation. The only relevant evidence comes from experiments on motor learning, the one area in which there are rather clear-cut operational definitions of what constitutes  $I_R$  and  $sI_R$ . In general, performance decrement that dissipates during rest is identified with  $I_R$ ; the decrement that still remains after rest is identified with  $sI_R$ .

Duncan (1951) gave two groups of subjects massed and distributed practice on the pursuit rotor. During this 5-minute practice period, the massed group presumably would develop more  $I_R$  and hence more  $sI_R$ . Then both groups were allowed 10 minutes of rest, so that at the beginning of the postrest trials, nearly all  $I_R$  should have dissipated, leaving the two groups differing only in  $sI_R$ . The postrest trials were massed for both groups. Here exist the very conditions which should allow an  $I_R \times sI_R$  interaction to show itself. If there were an interaction, the postrest performance curves of the two groups should diverge. In fact, they did not diverge, or converge, but ran exactly parallel throughout the postrest trials, which suggests an additive rather than multiplicative relationship between  $I_R$  and  $sI_R$ . There are certain weaknesses and peculiarities in Duncan's study (for example, it could be argued that the 5 minutes' practice was not sufficient to attain the threshold of  $I_R$  necessary for the development of  $sI_R$ , the evidence for which has been presented by Kimble, 1950); but on the whole, it favors Hull's formulation regarding inhibition more than it favors those formulations which involve  $I_R \times sI_R$ . Another study by Starkweather and Duncan (1954) was essentially the same as the previous experiment except that the massed

group was given more prerest practice so that performance on the first postrest trial would be the same for both massed and distributed groups. The rest period was 24 hours. Again, when both groups were given massed practice after the rest, their performance curves were approximately parallel, suggesting that there is no interaction between  $I_R$  and  $sI_R$ . It is possible to argue from some of the evidence in this study, however, that the presence of  $sI_R$  was not clearly demonstrated.<sup>4</sup>

Better evidence is presented by Bourne and Archer (1956). Groups trained under massed and distributed practice on the pursuit rotor were given 5 minutes' rest, and then all groups performed under massed conditions. The performance curves converged in the postrest period. But the convergence consisted of the performance of the previously distributed group reducing to that of the massed group. If the  $I_R \times sI_R$  formulation were correct, the result should have been just the opposite, with the previously massed group showing an increase up to the level of the distributed group. The prerest practice was more prolonged in this study than in Duncan's, and it can be argued that there was a sufficient amount of  $sI_R$  generated to permit the  $I_R \times sI_R$  to show itself. Yet, in another motor learning experiment specially designed to determine if there was an interaction between  $I_R$  and  $sI_R$ , Bowen, Ross, and Andrews (1956) failed to find any evidence of interaction. So while the evidence is not definitive on this point, the preponderance of it does not favor the

<sup>4</sup> It seems fairly certain that the concept of  $sI_R$  invoked to explain decremental phenomena in motor learning could not represent the same process as the  $sI_R$  involved in experimental extinction.

$I_R \times sI_R$  formulation. The issue, however, does not seem beyond a clear-cut experimental test. For example, in the Jones revision  $D \times sI_R$  would always have to be greater than  $I_R \times sI_R$ , because there can be no performance when  $D$  is equal to or less than  $I_R$ . If this were true, a person practicing on the pursuit rotor over a long period should finally become unable to perform, since  $sI_R$  would continue to grow and inhibit performance. After  $I_R$  had dissipated,  $D \times sI_R$  would approach or equal  $D \times sH_R$ , and the subject would be unable to perform the pursuit task. Gleitman, Nachmais, and Neisser (1954) were the first to point out this consequence with respect to Hull's formulation. As far as the writer knows, no one has ever found this kind of "extinction" of the pursuit rotor skill. Subjects have been known to practice the pursuit task day after day for months, long after having reached an asymptote for time on target, yet they show no loss of the skill. Hull's formula, on the other hand, can get around this problem, the arguments of Gleitman et al. (1954) notwithstanding. If  $sH_R$  and  $sI_R$  both reach an asymptote (Hull, 1951), extinction will have occurred when  $sI_R = D \times sH_R$ . An increase in  $D$  will make it possible for  $D \times sH_R$  to be greater than the asymptote of  $sI_R$ , so that extinction need never occur if  $D$  remains sufficiently high. Indeed, there are instances (Solomon & Wynne, 1954) of absence of extinction in escape and avoidance training in which the drive is a very strong shock-induced fear reaction.

The unlikely prediction made from Hull's theory by Gleitman et al. (1954) that any response, even though always reinforced, would eventually extinguish if it were repeated often enough was directly tested in experiments by Calvin, Clifford, Clifford,

Bolden, and Harvey (1956) and Kendrick (1958). Their studies differ in a few details of experimental procedure. Essentially they ran rats down a long alleyway at the end of which the rats received reinforcement *on every trial*. After some hundreds of trials (spread over many days) all the rats ceased running down the alley; they would not leave the starting box for a specified period of time designated as the criterion for "complete" extinction. Though this outcome lends support to Hull's theory, other interpretations are certainly possible (see Mowrer, 1960, pp. 426-432; Prokasy, 1960). The results of the Calvin et al. and Kendrick experiments may well be due to peculiarities of the experimental procedure. If not, one should expect "extinction with reinforcement" to occur in many other kinds of performance, such as a rat's bar pressing or a pigeon's pecking in a Skinner box, and in many types of repetitious motor tasks.

One experiment is highly relevant to theoretical predictions regarding the effects of drive on motor learning. Wasserman (1951), using a motor learning task (alphabet printing) found that high motivation resulted in performance which was significantly superior to that of low motivation (in both massed and distributed practice groups), the difference becoming progressively greater as practice continued. The Jones revision would predict just the opposite. Since  $D$  must always be greater than  $I_R$ ,  $D \times sI_R$  would result in greater performance decrement for the highly motivated group. The motivation in this experiment was controlled by the instructions given to the subjects, one group being task-oriented, the other ego-oriented.

#### $I_R \times sH_R$

This formulation of an interaction

between reactive inhibition and habit strength implies that the decremental effects on performance caused by the conditions producing  $I_R$  (effort and rate of response) will be greater for strong than for weak habits. This is patently incorrect, since it is known that there is a *positive* correlation between number of reinforced responses, of which  $sH_R$  is a function, and the number of responses emitted during extinction. The  $I_R \times sH_R$  formulation would predict just the opposite, i.e., a *negative* correlation between number of reinforcements and number of responses to extinction. This conclusion is not weakened by the fact (for example, Reid, 1953) that in learning to make a discrimination reversal the animals that have had a greater number of prereversal trials learn the reversal more quickly. This phenomenon may be interpreted in terms of the animal's also over-learning the *act* of making a discrimination (in addition to learning to respond differentially to the  $S^D$  and  $S^A$ ), which facilitates the learning of the reversal.

#### $sH_R \times sI_R$

This formulation, derived from Iwahara (1957), is subject to the same criticism just made in the case of  $I_R \times sH_R$ . It implies that the stronger the habit, the more quickly it should extinguish, which certainly is not true.

#### $K - I_R$

The suggestion of Woodworth and Schlosberg (1954), that total inhibition ( $I_R = I_R + sI_R$ ) be subtracted from incentive motivation,  $K$  (a function of amount of reinforcement), seems plausible, in that extinction involves the withdrawal of incentive. Within the total Hullian formulation, however, the Woodworth and Schlosberg suggestion meets with the same

difficulties pointed out in the two previous cases. Thus:

$$sE_R = D \times (K - I_R - sI_R) \times sH_R$$

In expanded form:

$$sE_R = D \times K - D \times I_R - D \times sI_R \\ \times D \times sH_R \times K \times sH_R \\ - sH_R \times I_R - sH_R \times sI_R$$

Thus we have again all of the elements that have already been criticized. Spence (1956) has argued, on the basis of experimental findings, that  $D$  and  $K$  are additive rather than multiplicative as in Hull. But here again the defects of the Woodworth and Schlosberg suggestion of  $K - I_R$  are evident.

$$sE_R = (D + K - I_R) \times sH_R$$

Expanded:

$$sE_R = D \times sH_R + K \times sH_R - sH_R \times I_R$$

The last term in the expanded formula again meets with the same difficulty pointed out above. It must be concluded that the  $K - I_R$  formulation is not an improvement on Hull or Spence.

#### SUMMARY

Several attempts to reformulate Hull's theory with respect to the inhibition postulates have been criticized. Because of the limitations of both Hull and his revisers in the exact quantification of intervening variables, much of the choice between alternative versions of the theory must be made on the basis of *plausibility* of congruence with empirical findings rather than of *prediction* of these findings in the rigorous sense of the term. All of the attempted revisions to date, with the possible exception of that of Spence, have serious shortcomings in the light of experimental evidence. They cannot, therefore, be regarded as improve-

ments over Hull's original formulation of reaction potential. Advances will be made, not by the mere algebraic manipulation of Hull's intervening variables—the method that

characterizes the present attempts—but by the postulation and quantification of new intervening variables, along with the experimental investigation of their interactions.

## REFERENCES

BARRY, H. Effects of strength of drive on learning and extinction. *J. exp. Psychol.*, 1958, 55, 473-481.

BASS, M. J., & HULL, C. L. The irradiation of tactile conditioned reflex in man. *J. comp. Psychol.*, 1934, 17, 47-65.

BITTERMAN, M. E., & HOLTZMAN, W. H. Conditioning and extinction of the galvanic skin response as a function of anxiety. *J. abnorm. soc. Psychol.*, 1952, 47, 615-623.

BOURNE, L. E., JR., & ARCHER, E. J. Time continuously on target as a function of distribution of practice. *J. exp. Psychol.*, 1956, 51, 25-33.

BOWEN, J. H., ROSS, S., & ANDREWS, T. G. A note on the interaction of conditioning and reactive inhibition in pursuit tracking. *J. gen. Psychol.*, 1956, 55, 153-162.

BRANDAUER, C. M. A confirmation of Webb's data concerning the action of irrelevant drivers. *J. exp. Psychol.*, 1953, 45, 150-152.

BROADHURST, P. L. Emotionality and the Yerkes-Dodson law. *J. exp. Psychol.*, 1957, 54, 345-352.

BROWN, JANET L. The effect of drive on learning with secondary reinforcement. *J. comp. physiol. Psychol.*, 1956, 49, 254-260.

BULLOCK, D. H. The inter-relationship of operant level, extinction ratio, and reserve. *J. exp. Psychol.*, 1950, 40, 802-804.

BULLOCK, D. H., & SMITH, W. C. An effect of repeated conditioning-extinction upon operant strength. *J. exp. Psychol.*, 1953, 46, 349-352.

CALVIN, A. A., CLIFFORD, T., CLIFFORD, B., BOLDEN, L., & HARVEY, J. An experimental validation of conditioned inhibition. *Psychol. Rep.*, 1956, 2, 51-56.

CAUTELA, J. R. Experimental extinction and drive during extinction in a discrimination habit. *J. exp. Psychol.*, 1956, 51, 299-302.

COTTON, J. W. On making predictions from Hull's theory. *Psychol. Rev.*, 1955, 67, 303-314.

CROCETTI, C. P. The relation of extinction responding to drive level in the white rat. Unpublished doctoral dissertation, Columbia University, 1952.

DINSMOOR, J. A. The effect of hunger on discriminated responding. *J. abnorm. soc. Psychol.*, 1952, 47, 67-72.

DUNCAN, C. P. The effect of unequal amounts of practice on motor learning before and after rest. *J. exp. Psychol.*, 1951, 42, 257-264.

ELLISON, D. G. The concept of reflex reserve. *Psychol. Rev.*, 1939, 46, 566-575.

EYSENCK, H. J. "Warm-up" in pursuit rotor learning as a function of the extinction of conditioned inhibition. *Acta Psychol., Amst.*, 1956, 12, 349-370.

EYSENCK, H. J. *The dynamics of anxiety and hysteria*. London: Routledge & Kegan Paul, 1957.

FITZWATER, M. E. The relative effect of reinforcement and nonreinforcement in establishing a form discrimination. *J. comp. physiol. Psychol.*, 1952, 45, 476-481.

GLEITMAN, H., NACHMAIS, J., & NEISER, U. The S-R reinforcement theory of extinction. *Psychol. Rev.*, 1954, 61, 23-33.

GRICE, G. R. The acquisition of a visual discrimination habit following response to a single stimulus. *J. exp. Psychol.*, 1948, 38, 633-642.

GRICE, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, 42, 365-373.

HANSON, H. M. Discrimination training effect on stimulus generalization gradient for spectrum stimuli. *Science*, 1957, 125, 888-889.

HILGARD, E. R. *Theories of learning*. (2nd ed.) New York: Appleton-Century-Crofts, 1956.

HILGARD, E. R. Intervening variables, hypothetical constructs, parameters, and constants. *Amer. J. Psychol.*, 1958, 71, 238-246.

HILGARD, E. R., JONES, L. V., & KAPLAN, S. J. Conditioned discrimination as related to anxiety. *J. exp. Psychol.*, 1951, 42, 94-99.

HILGARD, E. R., & MARQUIS, D. M. *Conditioning and learning*. New York: Appleton-Century-Crofts, 1940.

HOVLAND, C. I. The generalization of conditioned responses: The sensory generalization of conditioned responses with varying frequencies of tone. *J. gen. Psychol.*, 1937, 17, 125-148.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

HULL, C. L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951.

HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

IWAHARA, S. Hull's concept of inhibition: A revision. *Psychol. Rep.*, 1957, 3, 9-10.

JENKINS, W. O., & DAUGHERTY, GEORGETTE. Drive and the asymptote of extinction. *J. comp. physiol. Psychol.*, 1951, 44, 372-377.

JONES, H. G. The status of inhibition in Hull's system: A theoretical revision. *Psychol. Rev.*, 1958, 65, 179-182.

KENDRICK, D. C. Inhibition with reinforcement (conditioned inhibition). *J. exp. Psychol.*, 1958, 56, 313-318.

KIMBLE, G. A. Evidence for the role of motivation in determining the amount of reminiscence in pursuit rotor learning. *J. exp. Psychol.*, 1950, 40, 248-253.

KOCH, S., Clark L. Hull, In W. K. Estes, K. MacCorquodale, P. E. Meehl, C. G. Mueller, W. N. Schoenfeld, & W. S. Verplanck (Eds.), *Modern learning theory: A critical analysis of five examples*. New York: Appleton-Century-Crofts, 1954. Pp. 1-176.

LEWIS, D. J., & COTTON, J. W. Learning and performance as a function of drive strength during acquisition and extinction. *J. comp. physiol. Psychol.*, 1957, 50, 189-194.

LIBERMAN, A. M. A comparison of transfer effects during acquisition and extinction of two instrumental responses. *J. exp. Psychol.*, 1951, 41, 192-198.

MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, 55, 95-107.

MEYER, D. R. Food deprivation and discrimination reversal learning of monkeys. *J. exp. Psychol.*, 1951, 41, 10-16.

MILES, R. C. Discrimination in the squirrel monkey as a function of deprivation and problem difficulty. *J. exp. Psychol.*, 1959, 57, 15-19.

MOWRER, O. H. *Learning theory and behavior*. New York: Wiley, 1960.

OSGOOD, C. E. *Method and theory in experimental psychology*. New York: Oxford Univer. Press, 1953.

PAVLOV, I. P. *Conditioned reflexes*. London: Oxford Univer. Press, 1927.

PERIN, C. T. Behavior potentiality as a joint function of the amount of training and degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, 30, 93-113.

PERKINS, C. C., JR., & CACIOPPO, A. J. The effect of intermittent reinforcement on the change in extinction rate following successive reconditionings. *J. exp. Psychol.*, 1950, 40, 794-801.

PROKASY, W. F. Postasymptotic performance decrements during massed reinforcements. *Psychol. Bull.*, 1960, 57, 237-247.

RAZRAN, G. Extinction re-examined and re-analyzed: A new theory. *Psychol. Rev.*, 1956, 63, 39-52.

RAZRAN, G. H. S. Transposition of relational responses and generalization of conditioned responses. *Psychol. Rev.*, 1938, 45, 532-538.

REID, L. S. The development of noncontinuity behavior through continuity learning. *J. exp. Psychol.*, 1953, 46, 107-112.

REYNOLDS, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, 35, 15-30. (a)

REYNOLDS, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, 35, 81-95. (b)

REYNOLDS, B., MARX, M. H., & HENDERSON, R. L. Resistance to extinction as a function of drive-reward interaction. *J. comp. physiol. Psychol.*, 1952, 45, 36-42.

SACKETT, R. S. The effect of strength of drive at the time of extinction upon resistance to extinction in rats. *J. comp. Psychol.*, 1939, 27, 411-431.

SALTZMAN, I., & KOCH, S. The effect of low intensities of hunger on the behavior mediated by a habit of maximum strength. *J. exp. Psychol.*, 1948, 38, 347-370.

SHEFFIELD, VIRGINIA F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, 40, 305-313.

SHURRAGER, P. S., & SHURRAGER, H. C. Rate of learning measured at a single synapse. *J. exp. Psychol.*, 1946, 36, 347-354.

SKINNER, B. F. *The behavior of organisms*. New York: Appleton-Century-Crofts, 1938.

SMITH, M. H., & HAY, W. J. Rate of response during operant discriminations. *J. exp. Psychol.*, 1954, 48, 259-264.

SOLomon, R. L., & WYNNE, L. C. Traumatic avoidance learning: The principles of anxiety conservation and partial irreversibility. *Psychol. Rev.*, 1954, 61, 353-385.

SPENCE, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, 44, 430-444.

SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

SPENCE, K. W., & FARBER, I. E. Conditioning and extinction as a function of anxiety. *J. exp. Psychol.*, 1953, 45, 116-119.

SPENCE, K. W., & FARBER, I. E. The relation

of anxiety to differential eyelid conditioning. *J. exp. Psychol.*, 1954, 47, 127-134.

SPENCE, K. W., GOODRICH, K. P., & ROSS, L. E. Performance in differential conditioning and discrimination learning as a function of hunger and relative response frequency. *J. exp. Psychol.*, 1959, 58, 8-16.

STANLEY, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, 43, 246-260.

STARKWEATHER, J. A., & DUNCAN, C. P. A test for conditioned inhibition in motor learning. *J. exp. Psychol.*, 1954, 47, 351-356.

TEEL, K. S. Habit strength as a function of motivation during learning. *J. comp. physiol. Psychol.*, 1952, 45, 188-191.

WASSERMAN, H. N. The effect of motivation and amount of pre-rest practice upon inhibitory potential in motor learning. *J. exp. Psychol.*, 1951, 42, 162-172.

WEBB, W. B. A test of "relational" vs. "specific stimulus" learning in discrimination problems. *J. comp. physiol. Psychol.*, 1950, 43, 70-72.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. (Rev. ed.) New York: Holt, 1954.

YERKES, R. M., & DODSON, J. D. The relation of strength of stimulus to rapidity of habit formation. *J. comp. Neurol.*, 1908, 18, 459-482.

(Received February 11, 1960)

## ACQUIESCENCE AND THE FACTORIAL INTERPRETATION OF THE MMPI<sup>1</sup>

SAMUEL MESSICK  
*Educational Testing Service*

AND

DOUGLAS N. JACKSON  
*Pennsylvania State University*

The operation of reliable response sets or stylistic consistencies has been frequently noted on personality and attitude scales with a true-false or agree-disagree format (cf. Cronbach, 1946, 1950; Fricke, 1956; Messick & Jackson, 1958). It has recently been conjectured (Jackson & Messick, 1958) that the major common factors in personality inventories of this type are interpretable primarily in terms of such stylistic consistencies rather than in terms of specific item content. The present paper attempts to annotate the influence of two response styles, the tendency to agree or acquiesce and the tendency to respond in a desirable way, using the Minnesota Multiphasic Personality Inventory (MMPI) as an example of inventories with this general response form. In particular, a high correlation will be noted between factor loadings on the largest factor, as obtained in several published factor analyses of the MMPI, and certain indices of acquiescence.

Barnes (1956b), in evaluating the Berg (1955) deviation hypothesis on the MMPI, found that the tendency to answer atypically or deviantly "true" was highly correlated with

scores on the psychotic scales, and the tendency to answer atypically "false" was highly correlated with the neurotic triad. This result is consistent with the fact, noted by Cottle and Powell (1951) and others (Barnes, 1956b; Fricke, 1956), that a large proportion of MMPI psychotic items are keyed true and a large proportion of neurotic items keyed false, suggesting that differential tendencies to respond atypically "true" and "false" might have been involved in the discrimination of criterion groups upon which the scoring keys were based. Barnes (1956a) also pointed out a marked similarity between the correlations of MMPI scales with these two deviant response tendencies and factor loadings for the scales on the two major factors reported by Wheeler, Little, and Lehner (1951); he concluded that the number of atypical true answers is a "pure factor test" of the first or "psychotic" factor and that the number of deviant false answers has a high loading on the second or "neurotic" factor. The two major MMPI factors obtained by Welsh (1956) also displayed a similar pattern of loadings, and it is noteworthy that the "pure factor" reference scale *A* which Welsh developed for his first or "anxiety" factor had 38 out of 39 items keyed true, while the reference scale *R* for the second or "repression" factor had all 40 of its items keyed false.

In view of the striking similarity between the effects of consistent tendencies to respond "true" and "false" and patterns of factor loadings obtained in two studies of

<sup>1</sup> This study is part of a larger project on stylistic determinants in clinical personality assessment supported by the National Institute of Mental Health, United States Public Health Service, under Research Grants M-2878 to Educational Testing Service and M-2738 to Pennsylvania State University. The authors wish to thank George S. Welsh for graciously supplying scoring keys for his "pure" MMPI scales and Philip E. Slater for making available his factor analyses of the MMPI.

MMPI scales, all factor analyses of the MMPI readily available in the literature were reviewed, in order to evaluate the possible relationship between each scale's factor loading on the major factor and an index of its potential for reflecting acquiescence. The particular index of acquiescence used was the proportion of items keyed true on each scale, which, assuming that the acquiescence-evoking properties of items are uniform over all MMPI scales, can be considered to reflect the extent to which total scores on a scale are influenced by consistent tendencies to respond "true." High scores on a scale with a large proportion of items keyed true would thus be assumed to reflect a general tendency to acquiesce, in addition, of course, to the contribution of other stylistic tendencies and of systematic content responses. Jackson (1960) used this index to evaluate the effects of acquiescence on the California Psychological Inventory, and Voas (1958) used the proportion of items keyed false as a criterion for constructing response bias scales. Voas (1958) also estimated loadings for scales from the MMPI and the Guilford-Zimmerman Temperament Survey on a factor marked by two measures of the tendency to respond "false" and found that these loadings correlated .86 with the proportion of items keyed false on each scale. These findings support the use of the index in the present context.

Factor loadings for MMPI scales were obtained from eight studies by Abrams (1949, summarized by French, 1953), Cook and Wherry (1950), Cottle (1950), Tyler (1951), Wheeler, Little, and Lehner (1951), Welsh (1956), Slater (1958), and Kaszebaum, Couch, and Slater (1959). A fairly uniform finding from these

studies is that only two major factors and two or three minor ones are necessary to account for interrelations among the scales. Spearman rank correlations were computed between loadings on the largest factor in each study and the proportion of items keyed true on each scale; the results are summarized in Table 1. In some of the factor analyses, values were not reported for scales with small loadings on the factor, so in computing correlation coefficients these scales were considered to be tied at an appropriate rank below scales with reported positive loadings and above scales with reported negative loadings. Corrections for ties (cf. Siegel, 1956) were computed for two of the studies with the most scales tied at the same rank (Wheeler, Little, & Lehner's normal sample and Tyler's sample), but the coefficients changed only .01.

Of 11 different subject samples represented in these eight studies, significant correlations were obtained for 8 of them, four of the coefficients exceeding .85. These strikingly consistent findings indicate that in most of these studies the largest factor on the MMPI is interpretable in terms of acquiescence. In evaluating the few apparently inconsistent results, it is important to note that for Abrams's (1949) neurotic sample, the correlation with the largest factor was -.15, but with the second largest it was .52. Also, in Tyler's (1951) study the correlation with the largest rotated factor was .33, but with the unrotated first centroid it was .52,  $p < .05$ . These findings suggest that for those studies in which the correspondence between the proportion of items keyed true and the factor loadings was not close, the factor structures could have been rotated to produce a higher correlation. Ana-

TABLE 1

SPEARMAN RANK CORRELATION ( $\rho$ ) BETWEEN FACTOR LOADINGS ON THE LARGEST MMPI FACTOR AND PROPORTION OF ITEMS KEYED "TRUE" ON EACH SCALE

Study	Scales Included	Sample	$\rho$
Abrams, 1949	11 scales: <i>L, F, Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma</i>	117 normal male veterans	.907**
		201 neurotic male veterans	-.148 (largest factor) .516 (2nd largest)
Cook & Wherry, 1930	11 scales: <i>L, F, Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma</i>	111 male naval submarine candidates	.605*
Cottle, 1950	11 scales: <i>L, F, Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma</i>	400 male veterans	.916**
Tyler, 1951	15 scales: <i>Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma, Si, St, Pr, Ac, Re, Do</i>	107 female graduate students	.328
Wheeler, Little, & Lehner, 1951	12 scales: <i>L, K, F, Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma</i>	112 male college students	.558
		110 male neuropsychiatric patients	.874**
Welsh, 1936	11 pure scales: <i>K', Hs', D', Hy', Pd', Mf', Pa', Pt', Sc', Ma', Si'</i>	150 male VA general hospital patients	.870**
		Same 150 males	.897**
Slater, 1958	43 scales: <i>L, F, K, Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma, Si, Es, Nm, Dp, Fm, A, R, Im, Pr, To, C, P, Sp, Rp, Sy, Re, Si, Lp, Do, Es, Ie, Ac, Aj, O-I, Lb, Ns, Ca, Pt, Hs, Cm, Zi, Zs</i>	102 aged males	.728**
		109 aged females	.718**
Kassebaum, Couch, & Slater, 1959	32 scales: <i>L, F, K, Hs, D, Hy, Pd, Mf, Pa, Pt, Sc, Ma, Si, Es, Ie, Lp, Aj, Sy, Ac, Rs, Do, Pr, St, Im, Sp, Fm, Rp, R, A, Dp, To, OI</i>	160 Harvard College freshmen	.625**

\*  $p < .05$ .\*\*  $p < .01$ .

lytical procedures similar to the computation of B weights in multiple correlation analysis are available (Mosier, 1939) for rotating to maximize the correlation between a factor and a criterion, which in this case would be a vector of proportions of true items. However, an adequate application of this technique requires loadings for all the scales on the factors under consideration, and for those studies providing this information (e.g., Welsh, 1956) there was usually little need to rotate.

Another consideration which suggests that a rotation of axes might clarify the role of acquiescence on the MMPI is the fact that scales with high loadings on the second largest MMPI factor usually tend to have a high proportion of false items in their

keys. Kassebaum, Couch, and Slater (1959) noticed this in their factor results and suggested that their second factor partly reflected a general tendency to respond "false." Although correlations between the proportion of items keyed true and loadings on the second MMPI factor are usually not nearly as high as correlations with the first factor, some significant coefficients occur; e.g., the correlation between the proportion of items keyed true and loadings on the second factor in the study by Kassebaum, Couch, and Slater (1959) was  $-.44$ ,  $p < .05$  with 30  $df$ , and in Welsh's (1956) study it was  $-.64$ ,  $p < .05$  with 13  $df$ .

This result is consistent with Barnes' (1956a) finding of a correspondence between atypical true

answers and the first MMPI factor and atypical false answers and the second factor. Since these two factors are usually orthogonal, this correspondence might be considered evidence for two relatively independent response biases, one a tendency to agree and the other to disagree. Such a contention is consistent with Barnes' (1956b) finding of a correlation of .11 between deviant responses answered "true" and "false" and with the fact that Welsh's (1956) *A* and *R* scales are usually only slightly negatively correlated. Although these results cannot be accounted for by a simple response set of acquiescence, it is not necessary to postulate two independent sets to agree and to disagree. As has been pointed out (Jackson & Messick, 1958), all that is required to account for the findings is the operation of at least one other factor in conjunction with acquiescence. Thus, the *A* scale can have a high positive loading on an acquiescence factor and the *R* scale a high negative loading, yet the two scales could be uncorrelated if they both had positive, or negative, loadings on some other dimension. Other factors which could moderate the operation of acquiescence on the MMPI might be specific content dimensions or some other response style. As previously suggested (Jackson & Messick, 1958), a particularly likely candidate for such a role is the stylistic tendency to respond in a desirable way.

Possible influences on MMPI scores of a set to respond desirably have been widely documented (cf. De Soto & Kuehne, 1959; Edwards, 1957; Fordyce, 1956; Hanley, 1956, 1957; Jackson & Messick, 1958; Taylor, 1959; Wiggins & Rumrill, 1959). Fordyce (1956), for example, has noted a marked similarity between

loadings on the largest MMPI factor from Wheeler, Little, and Lehner's (1951) psychiatric sample and correlations of MMPI scales with a measure of desirability. In fact, the rank correlation between the loadings and the correlation coefficients is approximately  $-.75$ , and since the proportion of items keyed true on each MMPI scale correlates only about  $-.50$  with the desirability coefficients, it seems likely that a combination of desirability and acquiescence would lead to even better prediction of the factor (cf. Messick, 1959). Although this and some other reported relationships are somewhat equivocal because the measures of desirability used were partially confounded with acquiescence, e.g., Edwards' *SD* scale and Hanley's *Ex* scale, high correlations have also been reported between MMPI scales and desirability measures having a balanced number of true and false items (Edwards, 1957; Hanley, 1957; Wiggins & Rumrill, 1959).

In an attempt to take these findings into account, it is suggested that the acquiescence-evoking properties of items are not, as assumed above, uniform over all scales, but that acquiescence is elicited differentially as a function, perhaps, of specific item content, of the clarity or ambiguity with which the content is stated, and in particular of the perceived desirability of the statement. In the extreme, it is suggested that the two major factors usually found for the MMPI may be rotated into positions interpretable as two response styles—the tendency to acquiesce and the tendency to respond desirably. The negative poles of these dimensions would be the tendencies to disagree and to respond undesirably, respectively. Response vari-

ance on MMPI scales would then be primarily a function of these two stylistic components in various weighted proportions. Studies including independent marker variables for the two styles are of course required to identify the factor positions. Much research is also needed into the precise nature of the set to respond desirably, particularly in view of three complicating results: (a) the finding of consistent individual differences in judgments of desirability (Messick, 1960); (b) the distinction between personal and social desirability (Borislow, 1958; Rosen, 1956); and (c) the differentiation between a

tendency to endorse certain desirable items which exhibit large mean shifts under desirability instructions and the tendency to endorse other desirable items which presumably reflect a group norm (Voas, 1958; Wiggins, 1959).

In conclusion, the findings offer clear evidence that acquiescence, as moderated by item desirability, plays a dominant role in personality inventories like the MMPI. Focused empirical investigations are required to develop a refined interpretation of these and other stylistic consistencies in terms of personality organization and psychopathology.

## REFERENCES

ABRAMS, E. N. A comparative factor analytic study of normal and neurotic veterans. Unpublished doctoral dissertation, University of Michigan, 1949.

BARNES, E. H. Factors, response bias, and the MMPI. *J. consult. Psychol.*, 1956, 20, 419-421. (a)

BARNES, E. H. Response bias and the MMPI. *J. consult. Psychol.*, 1956, 20, 371-374. (b)

BERG, I. A. Response bias and personality: The deviation hypothesis. *J. Psychol.*, 1955, 40, 61-72.

BORISLOW, B. The Edwards Personal Preference Schedule (EPPS) and fakability. *J. appl. Psychol.*, 1958, 42, 22-27.

COOK, E. B., & WHEERY, R. J. A factor analysis of MMPI and aptitude test data. *J. appl. Psychol.*, 1950, 34, 260-266.

COTILE, W. C. A factorial study of the Multi-phasic, Strong, Kuder, and Bell inventories using a population of adult males. *Psychometrika*, 1950, 15, 25-47.

COTILE, W. C., & POWELL, J. O. The effect of random answers to the MMPI. *Educ. psychol. Measmt.*, 1951, 11, 224-227.

CRONBACH, L. J. Response sets and test validity. *Educ. psychol. Measmt.*, 1946, 6, 475-494.

CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, 10, 3-31.

DE SOTO, C. B., & KUETHE, J. L. The set to claim undesirable symptoms in personality inventories. *J. consult. Psychol.*, 1959, 23, 496-500.

EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.

FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, 20, 171-175.

FRENCH, J. W. *The description of personality measurements in terms of rotated factors*. Princeton, N. J.: Educational Testing Service, 1953.

FRICKE, B. G. Response set as a suppressor variable in the OAIS and MMPI. *J. consult. Psychol.*, 1956, 20, 161-169.

HANLEY, C. Social desirability and responses to items from three MMPI scales: D, Sc, and K. *J. appl. Psychol.*, 1956, 40, 324-328.

HANLEY, C. Deriving a measure of test-taking defensiveness. *J. consult. Psychol.*, 1957, 21, 391-397.

JACKSON, D. N. Stylistic response determinants in the California Psychological Inventory. *Educ. psychol. Measmt.*, 1960, 20, 339-346.

JACKSON, D. N., & MESSICK, S. Content and style in personality assessment. *Psychol. Bull.*, 1958, 55, 243-252.

KASSEBAUM, G. G., COUCH, A. S., & SLATER, P. E. The factorial dimensions of the MMPI. *J. consult. Psychol.*, 1959, 23, 226-236.

MESSICK, S. Review of Allen Edwards', *The social desirability variable in personality assessment and research*. *Educ. psychol. Measmt.*, 1959, 19, 451-454.

MESSICK, S. Dimensions of social desirability.

*J. consult. Psychol.*, 1960, 24, 279-287.

MESSICK, S., & JACKSON, D. N. The measurement of authoritarian attitudes. *Educ. psychol. Measmt.*, 1958, 18, 241-253.

MOSIER, C. I. Determining a simple structure when loadings for certain tests are known. *Psychometrika*, 1939, 4, 149-162.

ROSEN, E. Self-appraisal, personal desirability, and perceived social desirability of personality traits. *J. abnorm. soc. Psychol.*, 1956, 52, 151-158.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SLATER, P. E. Personality structure in old age. Progress Report, 1958, Age Center of New England, Project M-1402, National Institute of Mental Health.

TAYLOR, J. B. Social desirability and MMPI performance: The individual case. *J. consult. Psychol.*, 1959, 23, 514-517.

TYLER, F. T. A factorial analysis of fifteen MMPI scales. *J. consult. Psychol.*, 1951, 15, 451-456.

VOAS, R. B. Relationships among three types of response sets. Report No. 15, 1958, Naval School of Aviation Medicine, Pensacola, Project NM 16 0111 Subtask 1.

WELSH, G. S. Factor dimensions *A* and *R*. In G. S. Welsh & W. G. Dahlstrom (Eds.), *Basic readings on the MMPI in psychology and medicine*. Minneapolis: Univer. Minnesota Press, 1956.

WHEELER, W. M., LITTLE, K. B., & LEHNER, G. F. J. The internal structure of the MMPI. *J. consult. Psychol.*, 1951, 15, 134-141.

WIGGINS, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *J. consult. Psychol.*, 1959, 23, 419-427.

WIGGINS, J. S., & RUMRILL, C. Social desirability in the MMPI and Welsh's factor scales *A* and *R*. *J. consult. Psychol.*, 1959, 23, 100-106.

(Received April 7, 1960)

## SCALES AND STATISTICS: PARAMETRIC AND NONPARAMETRIC<sup>1</sup>

NORMAN H. ANDERSON

*University of California, Los Angeles*

The recent rise of interest in the use of nonparametric tests stems from two main sources. One is the concern about the use of parametric tests when the underlying assumptions are not met. The other is the problem of whether or not the measurement scale is suitable for application of parametric procedures. On both counts parametric tests are generally more in danger than nonparametric tests. Because of this, and because of a natural enthusiasm for a new technique, there has been a sometimes uncritical acceptance of nonparametric procedures. By now a certain degree of agreement concerning the more practical aspects involved in the choice of tests appears to have been reached. However, the measurement theoretical issue has been less clearly resolved. The principal purpose of this article is to discuss this latter issue further. For the sake of completeness, a brief overview of practical statistical considerations will also be included.

A few preliminary comments are needed in order to circumscribe the subsequent discussion. In the first place, it is assumed throughout that the data at hand arise from some sort of measuring scale which gives numerical results. This restriction is implicit in the proposal to compare parametric and nonparametric tests

since the former do not apply to strictly categorical data (but see Cochran, 1954). Second, parametric tests will mean tests of significance which assume equinormality, i.e., normality and some form of homogeneity of variance. For convenience, parametric test, *F* test, and analysis of variance will be used synonymously. Although this usage is not strictly correct, it should be noted that the *t* test and regression analysis may be considered as special applications of *F*. Nonparametric tests will refer to significance tests which make considerably weaker distributional assumptions as exemplified by rank order tests such as the Wilcoxon *T*, the Kruskal-Wallis *H*, and by the various median-type tests. Third, the main focus of the article is on tests of significance with a lesser emphasis on descriptive statistics. Problems of estimation are touched on only slightly although such problems are becoming increasingly important.

Finally, a word of caution is in order. It will be concluded that parametric procedures constitute the everyday tools of psychological statistics, but it should be realized that any area of investigation has its own statistical peculiarities and that general statements must always be adapted to the prevailing practical situation. In many cases, as in pilot work, for instance, or in situations in which data are cheap and plentiful, nonparametric tests, shortcut parametric tests (Tate & Clelland, 1957), or tests by visual inspection may well be the most efficient.

<sup>1</sup> An earlier version of this paper was presented at the April 1959 meetings of the Western Psychological Association. The author's thanks are due F. N. Jones and J. B. Sidowski for their helpful comments.

### PRACTICAL STATISTICAL CONSIDERATIONS

The three main points of comparison between parametric and nonparametric tests are significance level, power, and versatility. Most of the relevant considerations have been treated adequately by others and only a brief summary will be given here. For more detailed discussion, the articles of Cochran (1947), Savage (1957), Sawrey (1958), Gaito (1959), and Boneau (1960) are especially recommended.

*Significance level.* The effects of lack of equinormality on the significance level of parametric tests have received considerable study. The two handiest sources for the psychologist are Lindquist's (1953) citation of Norton's work, and the recent article of Boneau (1960) which summarizes much of the earlier work. The main conclusion of the various investigators is that lack of equinormality has remarkably little effect although two exceptions are noted: one-tailed tests and tests with considerably disparate cell  $n$ 's may be rather severely affected by unequal variances.<sup>2</sup>

A somewhat different source of perturbation of significance level should also be mentioned. An over-all test of several conditions may show that something is significant but will not localize the effects. As is well known, the common practice of  $t$  testing pairs of means tends to inflate the significance level even when the over-all  $F$  is significant. An

<sup>2</sup> The split-plot designs (e.g., Lindquist, 1953) commonly used for the analysis of repeated or correlated observations have been subject to some criticism (Cotton, 1959; Greenhouse & Geisser, 1959) because of the additional assumption of equal correlation which is made. However, tests are available which do not require this assumption (Cotton, 1959; Greenhouse & Geisser, 1959; Rao, 1952).

analogous inflation occurs with nonparametric tests. There are parametric multiple comparison procedures which are rigorously applicable in many such situations (Duncan, 1955; Federer, 1955) but analogous nonparametric techniques have as yet been developed in only a few cases.

*Power.* As Dixon and Massey (1957) note, rank order tests are nearly as powerful as parametric tests under equinormality. Consequently, there would seem to be no pressing reason in most investigations to use parametric techniques for reasons of power if an appropriate rank order test is available (but see Snedecor, 1956, p. 120). Of course, the loss of power involved in dichotomizing the data for a median-type test is considerable.

Although it might thus be argued that rank order tests should be generally used where applicable, it is to be suspected that such a practice would produce negative transfer to the use of the more incisive experimental designs which need parametric analyses. The logic and computing rules for the analysis of variance, however, follow a uniform pattern in all situations and thus provide maximal positive transfer from the simple to the more complex experiments.

There is also another aspect of power which needs mention. Not infrequently, it is possible to use existing data to get a rough idea of the chances of success in a further related experiment, or to estimate the  $N$  required for a given desired probability of success (Dixon & Massey, 1957, Ch. 14). Routine methods are available for these purposes when parametric statistics are employed but similar procedures are available only for certain nonparametric tests such as chi square.

*Versatility.* One of the most remarkable features of the analysis of variance is the breadth of its applicability, a point which has been emphasized by Gaito (1959). For present purposes, the ordinary factorial design will serve to exemplify the issue. Although factorial designs are widely employed, their uses in the investigation and control of minor variables have not been fully exploited. Thus, Feldt (1958) has noted the general superiority of the factorial design in matching or equating groups, an important problem which is but poorly handled in current research (Anderson, 1959). Similarly, the use of replications as a factor in the design makes it possible to test and partially control for drift or shift in apparatus, procedure, or subject population during the course of an experiment. In the same way, taking experimenters or stimulus materials as a factor allows tests which bear on the adequacy of standardization of the experimental procedures and on the generalizability of the results.

An analogous argument could be given for latin squares, largely rehabilitated by the work of Wilk and Kempthorne (1955), which are useful when subjects are given successive treatments; for orthogonal polynomials and trend tests for correlated scores (Grant, 1956) which give the most sensitive tests when the independent variable is scaled; as well as for the multivariate analysis of variance (Rao, 1952) which is applicable to correlated dependent variables measured on incommensurable scales.

The point to these examples and to the more extensive treatment by Gaito is straightforward. Their analysis is more or less routine when parametric procedures are used. However, they are handled inade-

quately or not at all by current nonparametric methods.

It thus seems fair to conclude that parametric tests constitute the standard tools of psychological statistics. In respect of significance level and power, one might claim a fairly even match. However, the versatility of parametric procedures is quite unmatched and this is decisive. Unless and until nonparametric tests are developed to the point where they meet the routine needs of the researcher as exemplified by the above designs, they cannot realistically be considered as competitors to parametric tests. Until that day, nonparametric tests may best be considered as useful minor techniques in the analysis of numerical data.

Too promiscuous a use of  $F$  is, of course, not to be condoned since there will be many situations in which the data are distributed quite wildly. Although there is no easy rule with which to draw the line, a frame of reference can be developed by studying the results of Norton (Linquist, 1953) and of Boneau (1960). It is also quite instructive to compare  $p$  values for parametric and nonparametric tests of the same data.

It may be worth noting that one of the reasons for the popularity of nonparametric tests is probably the current obsession with questions of statistical significance to the neglect of the often more important questions of design and power. Certainly some minimal degree of reliability is generally a necessary justification for asking others to spend time in assessing the importance of one's data. However, the question of statistical significance is only a first step, and a relatively minor one at that, in the over-all process of evaluating a set of results. To say that a result is statistically significant simply gives reasonable ground for believing that

some nonchance effect was obtained. The meaning of a nonchance effect rests on an assessment of the design of the investigation. Even with judicious design, however, phenomena are seldom pinned down in a single study so that the question of replicability in further work often arises also. The statistical aspects of these two questions are not without importance but tend to be neglected when too heavy an emphasis is placed on  $p$  values. As has been noted, it is the parametric procedures which are the more useful in both respects.

#### MEASUREMENT SCALE CONSIDERATIONS

The second and principal part of the article is concerned with the relations between types of measurement scales and statistical tests. For convenience, therefore, it will be assumed that lack of equinormality presents no serious problem. Since the  $F$  ratio remains constant with changes in unit or zero point of the measuring scale, we may ignore ratio scales and consider only ordinal and interval scales. These scales are defined following Stevens (1951). Briefly, an ordinal scale is one in which the events measured are, in some empirical sense, ordered in the same way as the arithmetic order of the numbers assigned to them. An interval scale has, in addition, an equality of unit over different parts of the scale. Stevens goes on to characterize scale types in terms of permissible transformations. For an ordinal scale, the permissible transformations are monotone since they leave rank order unchanged. For an interval scale, only the linear transformations are permissible since only these leave relative distance unchanged. Some workers (e.g.,

Coombs, 1952) have considered various scales which lie between the ordinal and interval scales. However, it will not be necessary to take this further refinement of the scale typology into account here.

As before, we suppose that we have a measuring scale which assigns numbers to events of a certain class. It is assumed that this measuring scale is an ordinal scale but not necessarily an interval scale. In order to fix ideas, consider the following example. Suppose that we are interested in studying attitude toward the church. Subjects are randomly assigned to two groups, one of which, reads Communication A, while the other reads Communication B. The subjects' attitudes towards the church are then measured by asking them to check a seven category pro-con rating scale. Our problem is whether the data give adequate reason to conclude that the two communications had different effects.

To ascertain whether the communications had different effects, some statistical test must be applied. In some cases, to be sure, the effects may be so strong that the test can be made by inspection. In most cases, however, some more objective method is necessary. An obvious procedure would be to assign the numbers 1 to 7, say, to the rating scale categories and apply the  $F$  test, at least if the data presented some semblance of equinormality. However, some writers on statistics (e.g., Siegel, 1956; Senders, 1958) would object to this on the ground that the rating scale is only an ordinal scale, the data are therefore not "truly numerical," and hence that the operations of addition and multiplication which are used in computing  $F$  cannot meaningfully be applied to the scores. There are three different

questions involved in this objection, and much of the controversy over scales and statistics has arisen from a failure to keep them separate. Accordingly, these three questions will be taken up in turn.

*Question 1. Can the F test be applied to data from an ordinal scale?* It is convenient to consider two cases of this question according as the assumption of equinormality is satisfied or not. Suppose first that equinormality obtains. The caveat against parametric statistics has been stated most explicitly by Siegel (1956) who says:

The conditions which must be satisfied . . . before any confidence can be placed in any probability statement obtained by the use of the *t* test are at least these: . . . 4. The variables involved must have been measured in *at least* an interval scale . . . (p. 19). (By permission, from *Nonparametric Statistics*, by S. Siegel. Copyright, 1956. McGraw-Hill Book Company, Inc.)

This statement of Siegel's is completely incorrect. This particular question admits of no doubt whatsoever. The *F* (or *t*) test may be applied without qualm. It will then answer the question which it was designed to answer: can we reasonably conclude that the difference between the means of the two groups is real rather than due to chance? The justification for using *F* is purely statistical and quite straightforward; there is no need to waste space on it here. The reader who has doubts on the matter should postpone them to the discussion of the two subsequent questions, or read the elegant and entertaining article by Lord (1953). As Lord points out, the statistical test can hardly be cognizant of the empirical meaning of the numbers with which it deals. Consequently, the validity of a statistical inference cannot depend on the type of measuring scale used.

The case in which equinormality does not hold remains to be considered. We may still use *F*, of course, and as has been seen in the first part, we would still have about the same significance level in most cases. The *F* test might have less power than a rank order test so that the latter might be preferable in this simple two group experiment. However, insofar as we wish to inquire into the reliability of the difference between the measured behavior of the two groups in our particular experiment, the choice of statistical test would be governed by purely statistical considerations and have nothing to do with scale type.

*Question 2. Will statistical results be invariant under change of scale?* The problem of invariance of result stems from the work of Stevens (1951) who observes that a statistic computed on data from a given scale will be invariant when the scale is changed according to any given permissible transformation. It is important to be precise about this usage of invariance. It means that if a statistic is computed from a set of scale values and this statistic is then transformed, the identical result will be obtained as when the separate scale values are transformed and the statistic is computed from these transformed scale values.

Now our scale of attitude toward the church is admittedly only an ordinal scale. Consequently, we would expect it to change in the direction of an interval scale in future work. Any such scale change would correspond to a monotone transformation of our original scale since only such transformations are permissible with an ordinal scale. Suppose then that a monotone transformation of the scale has been made subsequent to the experiment on attitude change.

We would then have two sets of data: the responses as measured on the original scale used in the experiment, and the transformed values of these responses as measured on the new, transformed scale. (Presumably, these transformed scale values would be the same as the subjects would have made had the new scale been used in the original experiment, although this will no doubt depend on the experimental basis of the new scale.) The question at issue then becomes whether the same significance results will be obtained from the two sets of data. If rank order tests are used, the same significance results will be found in either case because any permissible transformation leaves rank order unchanged. However, if parametric tests are employed, then different significance statements may be obtained from the two sets of data. It is possible to get a significant  $F$  from the original data and not from the transformed data, and vice versa. Worse yet, it is even logically possible that the means of the two groups will lie in reverse order on the two scales.

The state of affairs just described is clearly undesirable. If taken uncritically, it would constitute a strong argument for using only rank order tests on ordinal scale data and restricting the use of  $F$  to data obtained from interval scales. It is the purpose of this section to show that this conclusion is unwarranted. The basis of the argument is that the naming of the scales has begged the psychological question.

Consider interval scales first, and imagine that two students, P and Q, in an elementary lab course are assigned to investigate some process. This process might be a ball rolling on a plane, a rat running an alley, or a child doing sums. The students

cooperate in the experimental work, making the same observations, except that they use different measuring scales. P decides to measure time intervals. He reasons that it makes sense to speak of one time interval as being twice another, that time intervals therefore form a ratio scale, and hence a fortiori an interval scale. Q decides to measure the speed of the process (feet per second, problems per minute). By the same reasoning as used by P, Q concludes that he has an interval scale also. Both P and Q are aware of current strictures about

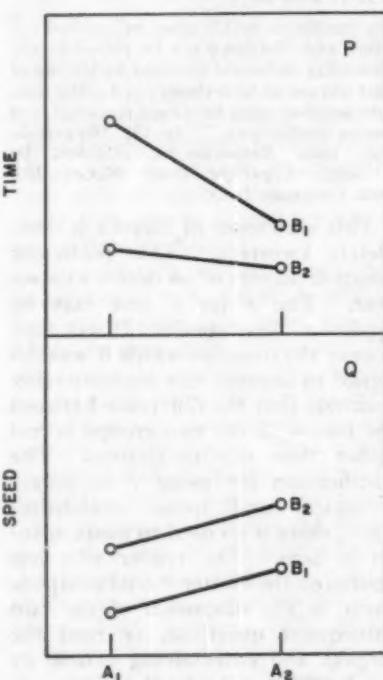


FIG. 1. Temporal aspects of some process obtained from a  $2 \times 2$  design. (The data are plotted as a function of Variable A with Variable B as a parameter. Subscripts denote the two levels of each variable. Note that Panel P shows an interaction, but that Panel Q does not.)

scales and statistics. However, since each believes (and rightly so) that he has an interval scale, each uses means and applies parametric tests in writing his lab report. Nevertheless, when they compare their reports they find considerable difference in their descriptive statistics and graphs (Figure 1), and in their *F* ratios as well. Consultation with a statistician shows that these differences are direct consequences of the difference in the measuring scales. Evidently then, possession of an interval scale does not guarantee invariance of interval scale statistics.

For ordinal scales, we would expect to obtain invariance of result by using ordinal scale statistics such as the median (Stevens, 1951). Let us suppose that some future investigator finds that attitude toward the church is multidimensional in nature and has, in fact, obtained interval scales for each of the dimensions. In some of his work he chanced to use our original ordinal scale so that he was able to find the relation between this ordinal scale and the multidimensional representation of the attitude. His results are shown in Figure 2. Our ordinal scale is represented by the curved line in the plane of the two dimensions. Thus, a greater distance from the origin as measured along the line stands for a higher value on our ordinal scale. Points A and B on the curve represent the medians of Groups A and B in our experiment, and it is seen that Group A is more pro-church than Group B on our ordinal scale. The median scores for these two groups on the two dimensions are obtained simply by projecting Points A and B onto the two dimensions. All is well on Dimension 2 since there Group A is greater than Group B. On Dimension 1, however, a reversal is found: Group A is less than Group B,

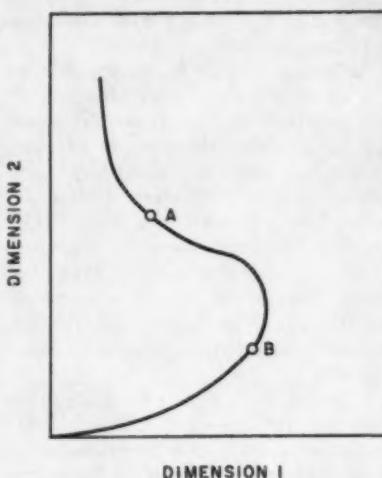


FIG. 2. The curved line represents the ordinal scale of attitude toward the church plotted in the two-dimensional space underlying the attitude. (Points A and B denote the medians of two experimental groups. The graph is hypothetical, of course.)

contrary to our ordinal scale results. Evidently then, possession of an ordinal scale does not guarantee invariance of ordinal scale statistics.

A rather more drastic loss of invariance would occur if the ordinal scale were measuring the resultant effect of two or more underlying processes. This could happen, for instance, in the study of approach-avoidance conflict, or ambivalent behavior, as might be the case with attitude toward the church. In such situations, two people could give identical responses on the one-dimensional scale and yet be quite different as regards the two underlying processes. For instance, the same resultant could occur with two equal opposing tendencies of any given strength. Representing such data in the space formed by the underlying dimensions would yield a smear of points over an entire

region rather than a simple curve as in Figure 2.

Although it may be reasonable to think that simple sensory phenomena are one-dimensional, it would seem that a considerable number of psychological variables must be conceived of as multidimensional in nature as, for instance, with "IQ" and other personality variables. Accordingly, as the two cited examples show, there is no logical guarantee that the use of ordinal scale statistics will yield invariant results under scale changes.

It is simple to construct analogous examples for nominal scales. However, their only relevance would be to show that a reduction of all results to categorical data does not avoid the difficulty with invariance.

It will be objected, of course, that the argument of the examples has violated the initial assumption that only "permissible" transformations would be used in changing the measuring scales. Thus, speed and time are not linearly related, but rather the one is a reciprocal transformation of the other. Similarly, Dimension 1 of Figure 2 is no monotone transformation of the original ordinal scale. This objection is correct, to be sure, but it simply shows that the problem of invariance of result with which one is actually faced in science has no particular connection with the invariance of "permissible" statistics. The examples which have been cited show that knowing the scale type, as determined by the commonly accepted criteria, does not imply that future scales measuring the same phenomena will be "permissible" transformations of the original scale. Hence the use of "permissible" statistics, although guaranteeing invariance of result over the class of "permissible" transformations, says little about

invariance of result over the class of scale changes which must actually be considered by the investigator in his work.

This point is no doubt pretty obvious, and it should not be thought that those who have taken up the scale-type ideas are unaware of the problem. Stevens, at least, seems to appreciate the difficulty when, in the concluding section of his 1951 article, he distinguishes between psychological dimensions and indicants. The former may be considered as intervening variables whereas the latter are effects or correlates of these variables. However, it is evident that an indicant may be an interval scale in the customary sense and yet bear a complicated relation to the underlying psychological dimensions. In such cases, no procedure of descriptive or inferential statistics can guarantee invariance over the class of scale changes which may become necessary.

It should also be realized that only a partial list of practical problems of invariance has been considered. Effects on invariance of improvements in experimental technique would also have to be taken into account since such improvements would be expected to purify or change the dependent variable as well as decrease variability. There is, in addition, a problem of invariance over subject population. Most researches are based on some handy sample of subjects and leave more or less doubt about the generality of the results. Although this becomes in large part an extrastatistical problem (Wilk & Kempthorne, 1955), it is one which assumes added importance in view of Cronbach's (1957) emphasis on the interaction of experimental and subject variables. In the face of these assorted difficulties, it is not easy to see what utility the scale typology

has for the practical problems of the investigator.

The preceding remarks have been intended to put into broader perspective that sort of invariance which is involved in the use of permissible statistics. They do not, however, solve the immediate problem of whether to use rank order tests or  $F$  in case only permissible transformations need be considered. Although invariance under permissible scale transformations may be of relatively minor importance, there is no point in taking unnecessary risks without the possibility of compensation.

On this basis, one would perhaps expect to find the greatest use of rank order tests in the initial stages of inquiry since it is then that measuring scales will be poorest. However, it is in these initial stages that the possibly relevant variables are not well-known so that the stronger experimental designs, and hence parametric procedures, are most needed. Thus, it may well be most efficient to use parametric tests, balancing any risk due to possible permissible scale changes against the greater power and versatility of such tests. In the later stages of investigation, we would be generally more sure of the scales and the use of rank order procedures would waste information which the scales by then embody.

At the same time, it should be realized that even with a relatively crude scale such as the rating scale of attitude toward the church, the possible permissible transformations which are relevant to the present discussion are somewhat restricted. Since the  $F$  ratio is invariant under change of zero and unit, it is no restriction to assume that any transformed scale also runs from 1 to 7. This imposes a considerable limitation on the permissible scale transfor-

mations which must be considered. In addition, whatever psychological worth the original rating scale possesses will limit still further the transformations which will occur in practice.

Although rank order tests do possess some logical advantage over parametric tests when only permissible transformations are considered, this advantage is, in the writer's opinion, very slight in practice and does not begin to balance the greater versatility of parametric procedures. The problem is, however, an empirical one and it would seem that some historical analysis is needed to provide an objective frame of reference. To quote an after-lunch remark of K. MacCorquodale, "Measurement theory should be descriptive, not prescriptive, nor prescriptive." Such an inquiry could not fail to be fascinating because of the light it would throw on the actual progress of measurement in psychology. One investigation of this sort would probably be more useful than all the speculation which has been written on the topic of measurement.

*Question 3. Will the use of parametric as opposed to nonparametric statistics affect inferences about underlying psychological processes?* In a narrow sense, Question 3 is irrelevant to this article since the inferences in question are substantive, relating to psychological meaning, rather than formal, relating to data reliability. Nevertheless, it is appropriate to discuss the matter briefly in order to make explicit some of the considerations involved because they are often confused with problems arising under the two previous questions. With no pretense of covering all aspects of this question, the following two examples will at least touch some of the problems.

The first example concerns the two students, P and Q, mentioned above, who had used time and speed as dependent variables. We suppose that their experiment was based on a  $2 \times 2$  design and yielded means as plotted in Figure 1. This graph portrays main effects of both variables which are seen to be similar in nature in both panels. However, our principal concern is with the interaction which may be visualized as measuring the degree of nonparallelism of the two lines in either panel. Panel P shows an interaction. The reciprocals of these same data, plotted in Panel Q, show no interaction. It is thus evident in the example, and true in general, that interaction effects will depend strongly on the measuring scales used.

Assessing an interaction does not always cause trouble, of course. Had the lines in Panel P, say, crossed each other, it would not be likely that any change of scale would yield uncrossed lines. In many cases also, the scale used is sufficient for the purposes at hand and future scale changes need not be considered. Nevertheless, it is clear that a measure of caution will often be needed in making inferences from interaction to psychological process. If the investigator envisages the possibility of future changes in the scale, he should also realize that a present inference based on significant interaction may lose credibility in the light of the rescaled data.

It is certainly true that the interpretation of interactions has sometimes led to error. It may also be noted that the usual factorial design analysis is sometimes incongruent with the phenomena. In a  $2 \times 2$  design it might happen, for example, that three of the four cell means are equal. The usual analysis is not optimally sensitive to this one real difference since it is distributed over

three degrees of freedom. In such cases, there will often be other parametric tests involving specific comparisons (Snedecor, 1956) or multiple comparisons (Duncan, 1955) which are more appropriate. Occasionally also, an analysis of variance based on a multiplicative model (Williams, 1952) will be useful (Jones & Marcus, 1961). A judicious choice of test may be of great help in dissecting the results. However, the test only answers set questions concerning the reliability of the results; only the research worker can say which questions are appropriate and meaningful.

Inferences based on nonparametric tests of interaction would presumably be less sensitive to certain types of scale changes. However, caution would still be needed in the interpretation as has been seen in Question 2. The problem is largely academic, however, since few nonparametric tests of interaction exist.<sup>3</sup> It might be suggested that the question of interaction cannot arise when only the ordinal properties of the data are considered since the interaction involves a comparison of differences and such a comparison is illegitimate with ordinal data. To the extent that this suggestion is correct, a parametric test can be used to the same purposes equally well if not better; to the extent that it is not correct, nonparametric tests will waste information.

One final comment on the first example deserves emphasis. Since both time and speed are interval scales, it cannot be argued that the

<sup>3</sup> There is a nomenclatural difficulty here. Strictly speaking, nonparametric tests should be called more-or-less distribution free tests. For example, the Mood-Brown generalized median test (Mood, 1950) is distribution free, but is based on a parametric model of the same sort as in the analysis of variance. As noted in the introduction, the usual terminology is used in this article.

difficulty in interpretation arises because we had only ordinal scales.

The second example, suggested by J. Kaswan, is shown in Figure 3. The graph, which is hypothetical, plots amount of aggressiveness as a function of amount of stress. A glance at the graph leads immediately to the inference that some sort of threshold effect is present. Under increasing stress, the organism remains quiescent until the stress passes a certain threshold value, whereupon the organism leaps into full scale aggressive behavior.

Confidence in this interpretation is shaken when we stop to consider that the scales for stress and aggression may not be very good. Perhaps, when future work has given us improved scales, these same data would yield a quite different function such as a straight line.

One extreme position regarding the threshold effect would be to say that the scales give rank order information and no more. The threshold inference, or any inference based on characteristics of the curve shape other than the uniform upward trend, would then be completely disallowed. At the other extreme, there would be complete faith in the scales and all inferences based on curve shape, including the threshold effect, would be made without fear that they would be undermined by future changes in the scales. In practice, one would probably adopt a position between these two extremes, believing, with Mosteller (1958), that our scales generally have some degree of numerical information worked into them, and realizing that to consider only the rank order character of the data would be to ignore the information that gives the strongest hold on the behavior.

From this ill-defined middleground, inferences such as the threshold effect

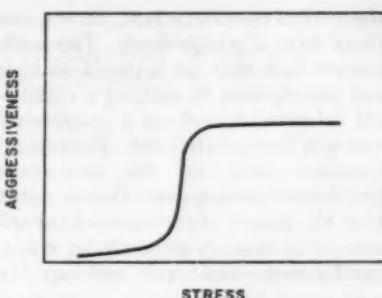


FIG. 3. Aggressiveness plotted as a function of stress. (The curve is hypothetical. Note the hypothetical threshold effect.)

would be entertained as guides to future work. Such inferences, however, are made at the judgment of the investigator. Statistical techniques may be helpful in evaluating the reliability of various features of the data, but only the investigator can endow them with psychological meaning.

#### SUMMARY

This article has compared parametric and nonparametric statistics under two general headings: practical statistical problems, and measurement theoretical considerations. The scope of the article is restricted to situations in which the dependent variable is numerical, thus excluding strictly categorical data.

Regarding practical problems, it was noted that the difference between parametric and rank order tests was not great insofar as significance level and power were concerned. However, only the versatility of parametric statistics meets the everyday needs of psychological research. It was concluded that parametric procedures are the standard tools of psychological statistics although nonparametric procedures are useful minor techniques.

Under the heading of measurement

theoretical considerations, three questions were distinguished. The well-known fact that an interval scale is not prerequisite to making a statistical inference based on a parametric test was first pointed out. The second question took up the important problem of invariance. It was noted that the practical problems of invariance or generality of result far transcend measurement scale typology. In

addition, the cited example of time and speed showed that interval scales of a given phenomenon are not unique. The discussion of the third question noted that the problem of psychological meaning is not basically a statistical matter. It was thus concluded that the type of measuring scale used had little relevance to the question of whether to use parametric or nonparametric tests.

## REFERENCES

ANDERSON, N. H. Education for research in psychology. *Amer. Psychologist*, 1959, 14, 695-696.

BONEAU, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, 57, 49-64.

COCHRAN, W. G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, 3, 22-38.

COCHRAN, W. G. Some methods for strengthening the common  $\chi^2$  tests. *Biometrics*, 1954, 10, 417-451.

COOMBS, C. H. A theory of psychological scaling. *Bull. Engrg. Res. Inst. U. Mich.*, 1952, No. 34.

COTTON, J. W. A re-examination of the repeated measurements problem. Paper read at American Statistical Association, Chicago, December 1959.

CRONBACH, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, 11, 671-684.

DIXON, W. J., & MASSEY, F. J., Jr. *Introduction to statistical analysis*. (2nd ed.) New York: McGraw-Hill, 1957.

DUNCAN, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, 11, 1-41.

FEDERER, W. T. *Experimental design*. New York: Macmillan, 1955.

FELDT, L. S. A comparison of the precision of three experimental designs employing a concomitant variable. *Psychometrika*, 1958, 23, 335-354.

GAITO, J. Nonparametric methods in psychological research. *Psychol. Rep.*, 1959, 5, 115-125.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141-154.

GREENHOUSE, S. W., & GEISSER, S. On methods in the analysis of profile data. *Psychometrika*, 1959, 24, 95-112.

JONES, F. N., & MARCUS, M. J. The subject effect in judgments of subjective magnitude. *J. exp. Psychol.*, 1961, 61, 40-44.

LINDQUIST, E. F. *Design and analysis of experiments*. Boston: Houghton Mifflin, 1953.

LORD, F. M. On the statistical treatment of football numbers. *Amer. Psychologist*, 1953, 8, 750-751.

MOOD, A. M. *Introduction to the theory of statistics*. New York: McGraw-Hill, 1950.

MOSTELLER, F. The mystery of the missing corpus. *Psychometrika*, 1958, 23, 279-290.

RAO, C. R. *Advanced statistical methods in biometrical research*. New York: Wiley, 1952.

SAVAGE, I. R. Nonparametric statistics. *J. Amer. Statist. Ass.*, 1957, 52, 331-344.

SAWREY, W. L. A distinction between exact and approximate nonparametric methods. *Psychometrika*, 1958, 23, 171-178.

SENDERS, V. L. *Measurement and statistics*. New York: Oxford, 1958.

SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

SNEDECOR, G. W. *Statistical methods*. (5th ed.) Ames: Iowa State Coll. Press, 1956.

STEVENS, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

TATE, M. W., & CLELLAND, R. C. *Nonparametric and shortcut statistics*. Danville, Ill.: Interstate, 1957.

WILK, M. B., & KEMPTHORNE, O. Fixed, mixed, and random models. *J. Amer. Statist. Ass.*, 1955, 50, 1144-1167.

WILLIAMS, E. J. The interpretation of interactions in factorial experiments. *Biometrika*, 1952, 39, 65-81.

(Received April 8, 1960)

## BASIC FORMS OF COVARIATION AND CONCOMITANCE DESIGNS

RICHARD W. COAN

*University of Arizona*

Several years ago, Cattell (1946) published a description of what he called the "covariation chart," a graphic model which illustrates six basic forms of covariation with which we may deal in psychological research. It is the purpose of the present paper to describe an extension and modification of Cattell's schema that will provide much more comprehensive classification of actual and possible research designs in psychology.

The six forms of covariation encompassed by Cattell's model have been variously labeled with letters through the alphabetic range from M to T, but the labeling indicated in Table 1 has come to be reasonably standard. The covariation chart itself consists of a parallelepiped, in which the three dimensions represent tests, persons, and occasions. Any plane parallel to any surface of the model represents a score matrix which might correspond to the data from a psychological research. There are three such sets of planes, any one

plane permitting consideration of two kinds of covariation.

The major virtue of a classification scheme like that embodied in the covariation chart is that it can suggest forms of valuable research which might otherwise be overlooked. As Cattell himself has clearly recognized, however, the scope of the covariation chart model has certain unfortunate limitations. When he first presented the covariation chart, Cattell pointed out that the six techniques did not really exhaust the forms of covariation inherently derivable from the three-dimensional model. The other forms which he considered at that time, however, are essentially variants or compounds of the six basic forms of covariation.

Various more novel techniques will emerge, of course, if we can find justification for adding other dimensions to the model. In a more recent publication, Cattell (1957) points out that a psychological event may be characterized in terms of six independent "tags": a reacting organism, a focal stimulus, a background condition, a response, an occasion in time and space, and an observer. He suggests that any pair of tags may serve as the dimensions of a score matrix yielding a technique and its transpose. Since there are 15 possible pairs of tags, there are 15 possible techniques (and their corresponding transposes). Furthermore, the elements within any matrix could correspond to any of the six tags. Logically, this would extend the system to 90 possible techniques (or 180, includ-

TABLE 1

THE SIX BASIC FORMS OF COVARIATION  
INDICATED IN THE COVARIATION CHART

Technique	Variables correlated	Series over which correlated	Variables held constant or singular
R	Tests	Persons	Occasions
Q	Persons	Tests	Occasions
P	Tests	Occasions	Persons
O	Occasions	Tests	Persons
S	Persons	Occasions	Tests
T	Occasions	Persons	Tests

ing transposed techniques). Cattell apparently excludes some combinations and speaks of 45 possible techniques. To these he adds five additional possibilities that involve a mixture of tags along one axis of the score matrix.

In the view of the writer, the original covariation chart provides too limited a classification system. The extended model, however, introduces needless complexity and is subject to useful modification and simplification. Cattell's six tags represent six distinguishable aspects of any observed psychological event, but they do not, on that account, constitute six meaningfully distinguishable aspects of research design.

The distinction between focal stimulus and background condition is a somewhat arbitrary one, and its usefulness in design classification is questionable. We can nearly always isolate a great variety of stimulus variables that will influence a given event in a more or less direct way. Insofar as the researcher analyzes the effect of one of these variables, it becomes a focal stimulus variable, at least from the standpoint of the researcher and hence of the research design. Background conditions are otherwise irrelevant to experimental design, unless they are confounded with other kinds of variables (organisms or occasions).

The observer is also a vital part of any psychological event dealt with in research, but the observer becomes important as a component of design only to the extent that he is something more than an observer. If his presence in the situation affects the behavior of the subject of the experiment, the observer becomes to that extent a part of the stimulus situation and may be analyzed accordingly. If our interest, on the other hand, is in

peculiarities of the observer as a recorder or rater of behavior, we are to that extent treating him as a reacting organism, i.e., as the subject of an experiment superimposed on another experiment.

#### BASIC COMPONENTS OF DESIGN IN PSYCHOLOGICAL RESEARCH

It is possible to characterize a psychological event in terms of a great number of distinguishable features which set it apart from other psychological events, but there are basically only four such features that constitute essential and distinguishable parameters of any research design employed to study psychological events. We shall refer to these features henceforth as *design components* and label them *R*, *S*, *P*, and *O* (not to be confused with *Techniques R*, *S*, *P*, and *O*).

Design Component *R* is that realm of variables which consists of structural or functional manifestations on the part of the subject or subjects under investigation and which are studied through observation and measurement of the subject or of products of the subject's behavior. Commonly treated as single *R*-component variables are specific responses, score summaries of patterns or sets of responses, and attributes. Design Component *S* is that realm of variables which arises from sources outside the subject and which may be expected to influence the subject's behavior. *S*, then, refers to external stimuli. Those things which are sometimes called "internal stimuli" fall within the scope of *R*-component variables if they are directly observed or measured. The *P* component is that of the human or animal subjects observed in the experiment. The *O* component is the realm of occasions, in given time and

space, on which experimental observations are made.

These four components are ordinarily quite distinct from one another and subject to separate specification. For some purposes, we may artificially tie variables of one component to those of another. In such cases, we may speak of a "confounding" of design components. Confounding is most common with respect to Component *O*, which for various purposes we permit to vary systematically with certain *S*, *P*, or *R* variables. A confounding of *S* and *P* variables is also quite common.

In a sense, any variable that we observe and describe may be said to be measured, at least implicitly, for if our description contains only an identifying qualitative statement, we have provided the essential ingredients of nominal scaling. Since the variables of all four design components are subject to observation and description in a psychological experiment, they may be regarded as subjected simultaneously and independently to measurement and scaling. Within any component, variables may be scaled at any level—nominal, ordinal, interval, or ratio—and are sometimes simultaneously scaled at more than one level.

A peculiarity of Component *P* that should be noted is that data within it are usually treated as scaled either at the nominal or at the ratio level. So long as we are concerned merely with identifying individuals as distinguishable entities, we make only the assumptions of nominal scaling. When we treat individuals as equivalent units that can be added together, however, and express *P*-component data in terms of numbers of cases or proportions of a total sample of subjects, we have made the essential assumptions underlying

ratio scaling. The data could be expressed in ordinal form if the label identifying the individual assumed the form of an index of rank within a social hierarchy. We could transform the data from ordinal form to presumably interval form either by making certain parametric assumptions or by adopting some appropriate measure of discriminability of adjacent ranks as an index of interval size. (Numerical data within the realm of Component *P* may assume any form consistent with the notion of *measurement in terms of individuals*. The application of measurement to individuals, however, yields *R*-component data.)

#### AN EXTENDED COVARIATION DESIGN CLASSIFICATION

A consideration of the role played by variables of the four design components in the covariation chart reveals that *R*-component variables are consistently assigned to the cells within the score matrices corresponding to Techniques *R*, *Q*, *P*, *O*, *S*, and *T*. The numbers in the body of a score matrix represent what we conceive of as the dependent variable in an experiment. In psychological research, the dependent variable is customarily, but not inevitably, the response variable. While our interest may lie in finding what sort of response will appear in a given situation, we may seek, with equal justification, to determine which individual will give a particular response, which stimulus will evoke the response, or on what occasion the response will appear. If we thus permit any of the four design components to furnish the elements within the score matrix, we are led to the system of 24 techniques shown in Table 2.

It may be noted that no component appears twice in any row of Table 2.

TABLE 2  
AN EXTENDED SYSTEM OF CO-VARIATION DESIGNS

Technique	Variables correlated	Variable in which variation is noted	Series over which covariation is studied	Constant or singular variable
R	S	R	P	O
Q	P	R	S	O
P	S	R	O	P
O	O	R	S	P
S	P	R	O	S
T	O	R	P	S
A	R	S	P	O
B	P	S	R	O
C	R	S	O	P
D	O	S	R	P
E	P	S	O	R
F	O	S	P	R
G	R	P	S	O
H	S	P	R	O
I	R	P	O	S
J	O	P	R	S
K	S	P	O	R
L	O	P	S	R
U	R	O	S	P
V	S	O	R	P
W	R	O	P	S
X	P	O	R	S
Y	S	O	P	R
Z	P	O	S	R

Note.—The letters in the second, third, fourth, and fifth columns refer to the design components from which variables are drawn.

This classification system assumes that the two axes of the matrix and the elements within the matrix will generally represent three different design components. Supporting this assumption is the fact that each design component represents variables which are an integral part of any psychological event, and the questions raised in psychological research normally refer to the manner in which variables of the different realms represented by the four components converge in a given psychological event. It must be granted, however, that our assumption is, in

some respects, an arbitrary one. It is possible to conceive of designs in which the axes and the matrix elements would not represent three different components, but such designs can also be rationalized quite readily as variants of techniques already in the system. Whether the classification system proposed here will generally provide the most convenient framework for design conceptualization must ultimately be determined through practical application. In any case, a classification system of this sort cannot be exhaustive if it is to remain fairly simple. It can merely provide a framework of basic prototypical techniques. Some designs will inevitably appear as combinations or variants of these techniques.

It must be emphasized that these techniques refer to research designs in which covariation is to be observed, but they do not imply any particular form of statistical analysis. In general, the desired indices of covariation will be furnished by correlational methods. Whether a method such as factor analysis or cluster analysis will be applied subsequently is an additional consideration.

#### COVARIATION DESIGN AND CONCOMITANCE DESIGN

If we are interested in truly comprehensive classification of psychological research designs, we must recognize at the outset that most psychological experiments are not actually concerned with covariation. The simplest form of research would call for a single measurement. This measurement might fall within the realm of any of our four design components, and it could be thought of as the single element filling a single-cell matrix. The variables of the other three components would also be singular.

More commonly we speak of re-

search design when we seek data for a matrix of at least two cells and where we are interested in a relationship among the ingredients of the matrix. The relationship may nearly always be considered in terms of a concomitance of two or more elements falling within the realm of one of our design components, and these elements are related in terms of their convergence with elements corresponding to a different component. If we represent all variables or elements of a common design component along a common axis of a score matrix, the data of many experiments must be thought of as filling cells arranged serially in a single row or column. We relate either the single cell rows of a single column matrix or the single cell columns of a single row matrix.

The kind of matrix we are now describing is a truncated version of the kind we assumed in classifying covariation designs. We can speak meaningfully of *concomitance* with respect to two single cell rows, but not of *covariation*, for this assumes two relatable series of values. In a single column matrix, whatever component would otherwise have constituted a horizontal axis is now treated as singular.

Nearly every psychological research design is concerned with concomitance, but not necessarily with covariation (i.e., concomitant variation). Since the covariation design is really a special case of concomitance design, it would be worthwhile to have a scheme of classification for concomitance designs which would parallel that for covariation designs. Such a scheme is presented in Table 3. Since in each concomitance design the serial variable is replaced by an additional singular variable, each concomitance design may be considered a truncated version of either of two covariation designs.

TABLE 3  
BASIC CONCOMITANCE DESIGNS

Technique	Variables related	Variables in which variation is noted	Singular or constant variables	Parallel covariation designs
Alpha	<i>S</i>	<i>R</i>	<i>P, O</i>	<i>R, P</i>
Beta	<i>P</i>	<i>R</i>	<i>S, O</i>	<i>Q, S</i>
Gamma	<i>O</i>	<i>R</i>	<i>S, P</i>	<i>O, T</i>
Delta	<i>R</i>	<i>S</i>	<i>P, O</i>	<i>A, C</i>
Epsilon	<i>P</i>	<i>S</i>	<i>R, O</i>	<i>B, E</i>
Zeta	<i>O</i>	<i>S</i>	<i>R, P</i>	<i>D, F</i>
Eta	<i>R</i>	<i>P</i>	<i>S, O</i>	<i>G, I</i>
Theta	<i>S</i>	<i>P</i>	<i>R, O</i>	<i>H, K</i>
Iota	<i>O</i>	<i>P</i>	<i>R, S</i>	<i>J, L</i>
Kappa	<i>R</i>	<i>O</i>	<i>S, P</i>	<i>U, W</i>
Lambda	<i>S</i>	<i>O</i>	<i>R, P</i>	<i>V, Y</i>
Mu	<i>P</i>	<i>O</i>	<i>R, S</i>	<i>X, Z</i>

Note.—The letters in the second, third, and fourth columns refer to design components from which variables are drawn.

#### APPLICATIONS OF CONCOMITANCE DESIGNS

The techniques labeled Alpha, Beta, and Gamma in Table 3 represent the most familiar forms of psychological research, and in them we find the most frequent application of such forms of statistical analysis as the critical ratio and analysis of variance. Beta technique has a common application in the comparison of responses of groups which differ with respect to variables outside the range of observation within the experiment (e.g., two different occupational groups, psychotics and "normals," men and women, etc.). Comparison of matched groups subjected to different stimulus conditions would constitute a form of Alpha technique, since *P*-component variables are held constant. Interest is here focused on the relating of stimuli, as in the simpler form of Alpha technique involving such a comparison for a single individual or single group of

individuals. A *compounding* of techniques is possible in designs of more than one-way classification. Thus, we should have a compound of Alpha and Beta techniques if we classified both in terms of known group membership and in terms of stimulus conditions. The reader will note that the score matrix in terms of which we conceptualize the design differs from the tabular arrangement usually employed with analysis of variance in that the variables to be related are represented along a common axis. Thus the score matrix for a complex factorial design of the Alpha-technique variety would consist of a long single column of  $R$  data. Each row would represent the data for a group simultaneously scaled with respect to several stimulus dimensions.

In Techniques Delta, Epsilon, and Zeta, the stimulus is conceptually the dependent variable. These techniques bring to mind certain applications of psychophysical methods. Strictly speaking, the procedures usually called "psychophysical methods," as described by such writers as Graham (1950) and Guilford (1954), are methods of measurement and do not define specific experimental designs to any greater extent than do methods of statistical analysis. In actual application, however, they form a basis for a limited range of concomitance designs.

The most common applications of psychophysical methods may be thought of as constituting either Alpha technique or Delta technique, depending largely on the use made of the data. The simple application of the method of constant stimuli, for example, would constitute Delta technique if we dealt with the resulting data in terms of a relationship between the two response categories. Each of the two cells of the corresponding score matrix would contain

the value of the stimulus eliciting the given response for a certain percentage of trials. On the other hand, findings may be expressed by means of a curve in which stimulus magnitude is plotted against the percentage of trials in which either response is produced. The design may then be considered either Alpha technique or Delta technique, depending on whether we consider the curve as a way of expressing relationships within a continuous series of  $S$  categories or within a continuous series of  $R$  categories (percentages in the present instance). Similar reasoning would apply, of course, to the application of other psychophysical methods. More complex applications of these methods, in which  $R$  variables are related to a combination of interacting  $S$  dimensions—as in Licklider's (1951) treatment of auditory functions—may be regarded as comparable to the application of factorial design in Alpha technique. Psychophysical methods are less commonly applied in research classifiable as Epsilon or Zeta technique, although certain applications of these methods in clinical research (e.g., certain studies involving flicker fusion, size judgments, distance judgments, and judgments of the vertical) would certainly qualify as Epsilon technique.

Techniques Eta, Theta, and Iota are a common realm of application for nonparametric techniques of statistical analysis. Depending on the manner in which  $P$ -component data are expressed, we may analyze findings in terms of cell frequencies, overlap of cases among cells, or comparability of person ranks associated with various cells.

Techniques Kappa, Lambda, and Mu are most likely to be useful when variations in an occasion variable are presumed to covary with certain

attributes of subjects or with certain changes in the life situations of subjects. The *O*-component variable may thus reflect such things as age, developmental stage, and level or stage of experience. The developmental area is probably the most common realm of application. Kappa technique provides a means for grouping behaviors developmentally and hence for defining developmental stages. Lambda technique provides a way of defining stages in terms of effective stimuli. Mu technique can be used to compare individuals with respect to such things as rate of maturation. Many applications outside the developmental realm, to processes involving shorter time spans, are possible.

#### APPLICATIONS OF COVARIATION DESIGNS

A detailed discussion of possible applications of the familiar Techniques *R*, *Q*, *P*, *O*, *S*, and *T* would be superfluous here. Unfortunately, other treatments of these techniques have promoted misconceptions by obscuring three interrelated considerations that are basic to consistent classification. First, there is the distinction between concomitance and covariation designs. A second vital point is that the series over which covariation is observed must be genuinely treated as a series in covariation designs. Wherever a group is treated as a unit and a group average is treated as a single observation, the group functions, for design classification purposes, as a single individual. Finally, in research employing matched groups, the *P* component is properly viewed as being held constant, and appropriate classification will depend on what component is confounded with Component *P*. For example, in the common type of experiment in which equated con-

trol and experimental groups are subjected to different stimulus conditions, we have an instance of *P* technique (not *S* technique, as some writers would have it), provided that response covariation is considered over time. The usual application of this design, to a single occasion, is simply Alpha technique.

The remaining techniques—*A* through *L* and *U* through *Z*—represent virtually unexploited forms of design, but careful consideration will suggest appropriate uses for each of them. In Techniques *A* through *F*, the dependent variable is of Component *S*. Appropriate quantification might be in terms of minimally sufficient stimulus magnitude or mean stimulus magnitude associated with a given response. In Techniques *G* through *L*, the dependent variable is of Component *P*. It may be expressed in terms of the rank of the individual giving a certain response to a certain stimulus on a certain occasion, in terms of the average rank of individuals so responding, or in terms of the number of individuals so responding. In Techniques *U*, *V*, *W*, *X*, *Y*, and *Z*, our focal variable—of Component *O*—may be expressed in terms of a single occasion in an ordered series, an average of a number of ordered occasions, an average age, an average stage, etc. It is important to note that in covariation designs, in contrast to concomitance designs, the dependent variable must be of at least the ordinal level of scaling. Thus, in Techniques Eta, Theta, and Iota, the matrix cells could simply contain tags identifying the persons fitting the cell coordinates. Data analysis would then consist of assessing the overlap of entries in various cells. In a matrix of several rows and columns containing such nominal data, we could probably speak of "multiple concomitance"

with respect to a pair of rows or columns, but it is doubtful that we can properly speak of "covariation" unless the data in the body of the matrix are expressed in a form representing relative magnitudes or positions on continua.

Techniques *A*, *C*, *G*, *I*, *U*, and *W* all deal with the covariation of response categories and thus provide a basis for defining the structure of a response realm. In *C* and *U*, we assess the comparability of response categories on an intra-individual basis, with respect to conjoint appearance on various occasions or in response to various stimuli. Techniques *A*, *G*, *I*, and *W* provide a means for assessing response comparability on a group basis, in terms of similarity of the precipitating stimulus, of the occasion of manifestation, or of the persons giving the response.

Techniques *R* and *P* are familiar techniques for examining stimulus covariation in terms of the resulting response. Techniques *H*, *K*, *V*, and *Y* add possibilities for correlating stimuli in terms of covariation with respect to the magnitudes (ranks) or numbers of persons responding in a certain way or in terms of the particular occasions or numbers of occasions on which the stimulus has a given effect. The correlating of persons is also a familiar idea by virtue of its introduction through *Q* and *S* techniques. Techniques *B*, *E*, *X*, and

*Z* point to the possibility of correlating persons according to stimuli producing various responses, stimuli producing a given response on various occasions, occasions when various responses appear, or occasions when various stimuli elicit a given response. Consideration of the many possible ways of defining the basic stimulus, response, and occasion data suggests a great variety of ways of grouping persons according to such things as physiological cycles, social roles, and developmental patterns.

In applying any of the occasion-correlation techniques—*O*, *T*, *D*, *F*, *J*, and *L*—we may select presumably equivalent occasions and thus obtain an estimate of the reliability, or stability, of a given pattern of relationship. We may, on the other hand, select occasions differing in a known way and thereby determine the comparability of these occasions. Possible applications range from the psychophysical realm to the developmental realm, depending on how the occasion variable is defined and quantified. In general, the new covariation techniques encompassed by this expanded classification system promise a rich harvest through novel approaches to diverse problems—particularly in the developmental, social, and physiological areas, where the possible fruits of correlational analysis have been recognized by too few researchers.

#### REFERENCES

CATTELL, R. B. *The description and measurement of personality*. New York: World Book, 1946.

CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.

GRAHAM, C. H. Behavior, perception, and the psychophysical methods. *Psychol. Rev.*, 1950, **57**, 108-120.

GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

LIKELIDER, J. C. R. Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 985-1039.

(Received April 28, 1960)

## THE SELF-CONCEPT: FACT OR ARTIFACT?

C. MARSHALL LOWE<sup>1</sup>

Ohio State University

One of the more difficult tasks for psychology is relating the observation of behavior to the study of mental processes. One approach to the problem has been to limit psychology to the study of behavior and to leave to philosophy the task of speculating as to the existence and nature of mind and soul.

There have, however, been psychologists who have sought to make sense out of human action by positing a self or ego, in order that they might understand the coherence and unity which they have thought that they have seen in human behavior. Thus, G. W. Allport (1943) claimed that the concept of ego was made necessary by certain shortcomings in associationism, and he went on to list eight different uses for the concept of the ego. During the 1940s the *Psychological Review* was in fact well-flavored with articles of philosophical taste (Allport, 1943; Bertocci, 1945; Chein, 1944; Lundholm, 1940). These articles were attempts to find the source of human behavior by discussions of concepts, but they failed to make a lasting distinction between the self as subjective knower and the self as object of knowledge. The self as essence defied definition, and the discussions concerning the nature of mind seemed relevant for neither experimental nor applied psychology.

But during the 1940s there was a parallel attempt at construction of a

useful concept of the self. While Rogers wrestled with the problem of researching a client centered approach in psychotherapy, one of his students (Raimy, 1943) developed a construct of the self which had a perceptual frame of reference. What Raimy called the self-concept was both a learned perceptual system functioning as an object in the perceptual field, and a complex organizing principle which schematizes ongoing experience. Raimy demonstrated in his dissertation that attitudes toward the self can be found by analyzing counseling protocols, and that these self-perceiving attitudes formed a reliable index for improvement in psychotherapy.

The concept of the self soon formed the theoretical underpinning for a new approach to the study of behavior. Raimy's construct of the self received further development in the book *Individual Behavior* (Snygg & Combs, 1949). The authors stated that behavior was best understood as growing out of the individual subject's frame of reference. Behavior was to be interpreted according to the phenomenal field of the subject rather than be seen in terms of the analytical categories of the observer.

As the self-concept was born with client centered therapy, so congruent were the theory of the self and the practice of psychotherapy that a new self centered therapy became theoretical for the first time: Rogers (1951) described therapeutic change in a phenomenological frame of reference.

<sup>1</sup> Now at Maryville College.

The author is indebted to Collins W. Burnett of the Department of Psychology for his encouragement and suggestions in the preparation of this manuscript.

By 1950 the phenomenological view of the self had become the center of a new movement in psychology, having already generated a block of research studies (Rogers et al., 1949). When Hilgard (1949) postulated in his APA presidential address the need for a self to understand psychoanalytic defense mechanisms, and called for research on the self, psychology listened. To the desert came rain that washed all before it.

The deluge of studies within the last decade has not been contained within any one theoretical channel, so that studies involving the self-concept have spread over into many areas of psychology. Ten years of research efforts have produced a mass of data, reflecting different theoretical assumptions and differing research methods. While the time has now passed for one article to deal adequately with all the studies that have been done, the sheer mass of evidence would suggest that certain questions be asked of theories of the self-concept.

This paper is concerned with the problem as to whether the self is an objective reality which is a fit field for psychological research, or whether it is a somewhat nebulous abstraction useful only to give a theoretical basis to things the psychologist could not otherwise understand. Put in other words, this paper faces the issue as to whether the results of studies of the self are to be accepted at face value, or whether other explanations of results would be more parsimonious or reasonable.

The writer will discuss first attempts to quantify data concerning the self-concept to arrive at an operational definition. We will then assess the validity of measures of the self-concept, and will relate the self-

concept to other constructs. We will briefly allude to attempts to establish a relationship between different measures. Finally, the writer will return to certain philosophical and historical considerations in order to reach a conclusion as to whether the self-concept is indeed a fact of nature, or an artifact of men's minds.

#### MEASURING THE SELF-CONCEPT

Many psychologists have believed that if something exists it can be measured. There have been many investigators who have assumed that the self-concept refers to an existence of some sort and have gone on to measure it.

The most popular type of operational definition has assumed that the self-concept can be defined in terms of the attitudes toward the self, as determined either by the subject's references to himself in psychotherapy or by asking him to mark off certain self-regarding attitudes on a rating scale.

One of the first attempts at attitude measurement was by Sheerer (1949), who extracted from the protocols of cases at the University of Chicago Counseling Center all statements that were relevant either for attitudes to self or to other people. These statements formed the basis for a 101-item rating scale. The Sheerer client statements also formed the basis for rating scales constructed by Phillips (1951) and by Berger (1952).

The only rating scale of attitudes towards self that has been published is the Index of Adjustment and Values (Bills, 1958). Bills states that the intent of the index is to measure the phenomenological self view as described by Lecky (1945), Snygg and Combs (1949), and Rogers (1951). This scale is more elaborate in that

each item is ranked with three different instructions. First, the subject ranks the item on a scale as to how well it describes himself. Next, he marks the items as to how acceptant he is of his first, or self-rating of the item, and finally he rates the item as to the degree to which he aspires to be like that item.

The scoring of the Bills index also is more elaborate than that traditional for rating scales. There are in fact two different measures, neither one being simply a rating of items in absolute terms, as in the scales previously described. Bills' measures depend instead upon the differences between ratings made under different instructions. A measure of self-acceptance is provided by the degree of similarity between the way the subject sees himself as being, and the way he rates himself as accepting his self-ratings. A measure of self-ideal-self discrepancy is given by comparing the differences in ratings between the way the self is rated as being, and the way the self is rated as wishing to be.

Brownfain (1952) made still another adaptation in the use of the rating scale, deriving a measure of what he termed the stability of the self-concept. Subjects ranked themselves on 25 words and phrases, each describing a different area of personality adjustment. The measure is not of how sure the subject is of himself, but of how sure he is of what he thinks about himself; the subject is instructed to make the ratings twice, first with an optimistic frame of reference, and then with a pessimistic one. The degree of congruence between the two ratings is termed the degree of stability of the self-concept.

A different theoretical approach towards measurement of self-concept involves the use of *Q* technique

Stephenson (1953) describes how one's "inner experiences" can be translated into behavior by means of *Q* sort, through which the phenomenal field is translated into action. Using this method, two of Stephenson's students at the University of Chicago derived a conceptual self-system in an intensive study of a single subject (Edelson & Jones, 1954).

Others at the University of Chicago have used *Q* sorts as a measure of self-concept, in an attempt to assess changes in self-concept during psychotherapy (Rogers & Dymond, 1954). Statements were taken from counseling protocols, and were sorted both for real self and for ideal self. The degree of congruence between the two sorts is taken as a measure of adjustment.

Attempts to measure the self-concept face three difficulties. First, it must be demonstrated that the operational and philosophic meanings are in fact equivalent. In the case of the self-concept it needs to be shown that the "inner experience" is effectively conveyed by the outward movement of making check marks on lines, or sorting cards. Secondly, an efficient and systematic method must be found for selecting items for the scales and sorts, the problem being that of defining the universe from which items are to be selected. Finally, the different measures imply different operational definitions. Just as one can not multiply apples and pears, so is it impossible to interchange different operational definitions as if they were the same, or to pretend that each means the same thing by the term self-concept.

If something is measured does it exist? If the answer is yes, we must still be aware that we may not fully understand what we are measuring.

One must measure, but must then compare and carefully validate.

#### VALIDATION OF SELF-CONCEPT MEASURES

A psychological construct stands and falls according to how useful it is in understanding human behavior. A term is meaningful only when successful validation studies have found significant relationships with established variables.

It has been popular to validate self-concept scales against tests purporting to measure maladjustment in an attempt to demonstrate that one's phenomenological view of the self is closely related to the degree of adjustment. Positive results abound. Calvin and Holtzman (1953) had college students rank themselves on seven personality traits, and found that self-depreciation was related to high scores on the MMPI. Zuckerman and Manashkin (1957) had neuropsychiatric patients rate themselves on a scale of adjectives, and found that self-ratings correlated positively with the MMPI *K* scale, and negatively with seven of the other scales. Taylor and Combs (1952) tested the hypothesis that sixth grade children found to be well-adjusted on the California personality scale would more often admit statements of self-reference which though unflattering were universally true. They got positive results, the self-depreciation which in other self-concept measures is treated as vice being here treated as virtue. Hanlon, Hofstaetter, and O'Connor (1954) compared the results of high school juniors on the California personality scale with the degree of congruence between ratings of the real and ideal self and found that the more congruence the better the adjustment. Cowen (1954) related low self-ratings

on the Brownfain negative self-concept with high scores on the California *F* Scale. Any doubt about the ability of investigators to find positive results when comparing good adjustment as measured by objective personality inventories with the affirmativeness of self-concept should be dispelled by a study by Smith (1958). He compared congruence between *Q* sorts for self and ideal self with scores on the Edwards PPS, the Cattell factors, and measures of average mood. After making almost 300 correlations, he concluded that having a positive self-concept is indeed related to adjustment.

Other investigators have doubted that the relationship between adjustment and self-satisfaction is such a simple one. Block and Thomas (1955) conceived of maladjustment lying at both ends of the continuum. They felt that too high a degree of self-satisfaction is due to suppressive and repressive mechanisms which cause a person to be rigid, over-controlled, restrained, and aloof. But at the other extreme, the person who is too little satisfied with self will lack ego defenses, and will be able neither to bind tensions nor control emotions. Block and Thomas constructed an ego-control scale from MMPI items. The scale was found to have a correlation of .44 with self-ideal-self *Q* sort congruence, the relationship being curvilinear. Unfortunately, this was the reverse of what Chodorkoff (1954a) had found. Correlating ratings of the self as made from a biographical inventory with the results of projective techniques, he found that maladjustment lies in the middle range of self-satisfaction.

Validating self-concept measures against objective personality tests has generally been successful, but the true significance of these studies is

still not made clear. Edwards (1957) demonstrates how more than half the variance in both MMPI scales and in *Q* sorts of self-referent items is accounted for by social desirability. SD can account for significant positive relationships even when other variables are totally unrelated. Edwards' SD robs these studies neither of significance nor of interest, but does suggest that extreme care must be taken in the labeling of constructs.

Attempts have also been made to validate self-concept against projective personality tests. Bills has made several attempts to validate his scale by the Rorschach (Bills, 1953a, 1954; Bills, Vance, & McLean, 1951). The results are a bit ambiguous, and leave two observers (Cowen & Tongas, 1959) extremely dissatisfied. The TAT was used by Friedman (1955) to compare the *Q* sort discrepancy self with the self as projected onto the TAT pictures. The normals were the only group to project positive self-qualities. Neurotics and paranoid both projected negatively.

A different approach to validation has used a word association test. Results show that there is a delayed reaction time for those trait words where there has been a discrepancy in ratings between the self and the ideal self (Bills, 1953b, Roberts, 1952). Delayed associations are assumed to be related to defensiveness about self, which in turn is considered to be related to maladjustment. However, Cowen and Tongas (1959) wonder if defensiveness about trait words does not serve also to raise the original ratings of the actual self.

Cowen chose to validate the self-concept by comparing the absolute self-rating with the learning time for the rated words, and found that

there was a higher learning time for words that were presumably threatening. We might however wonder if his and Tongas' criticism of other studies does not apply here also: defensiveness might also cause self-ratings to be raised.

Use was also made of the perceptual New Look. Chodorkoff (1954b) presented neutral and threatening words with a tachistoscope, and found that the better the agreement between a self-description and a description of the self by others, the less perceptual threat there will be.

It is unfortunate that the only study of this general type that did not use college students as subjects was negative in its results. Zimmer (1954) presented male mental patients with trait adjectives on which there was a self-rating discrepancy between self and ideal self. A word association test was not found to be significantly related to self-discrepancy.

The results of studies that involve the presentation of "hot" or threatening words seem suggestive, for there seems to be a common element in ability to free associate and learn threatening words. But it is possible that we have in these studies more a measure of ego defenses than of maladjustment, for the fact that the results are positive only with normal groups might suggest that the results are more relevant for a theory of personality than for a theory of psychopathology. In these studies it is indeed likely that we have support for Lecky's theory of self-consistency, and for Snygg and Comb's theory of the maintenance of internal organization. If this is so, then likely it is true as Block and Thomas (1955) suggest that only extremes in ego control are pathological.

A different approach to validation

of self-concept measures uses behavior in a social situation as a criterion. The most sweeping results in a study of this type are reported by Turner and Vanderlippe (1958), who report that *Q* sort congruence between the self and the ideal self is greater in those college students who are more active in extracurricular activities, have higher scholastic averages, and are given higher sociometric rankings by fellow students. Holt (1951) found that agreement between self-ratings and ratings by a diagnostic council was positively related to intelligent, active, adventurous living, and a friendly dominant social adjustment. Eastman (1958) found that the degree of acceptance of self-ratings on the Bills index is positively related to ratings for marital happiness. Working in terms of ratings for maladjustment, Chase (1957) found that among maladjusted patients there was greater discrepancy between *Q* sorts for self as compared with sorts for the ideal self and the average other person.

Other attempts to relate self-concept to social behavior have been less successful. Kelman and Parloff (1957) obtained only chance results when they tried to interrelate such variables as congruence between self and ideal self, a symptom disability check list, a discomfort evaluation scale, sociometric ratings, and an ineffective behavior evaluation scale, using 15 neurotic hospital outpatients. Fiedler, Dodge, Jones, and Hutchins (1958) measured the self-concept of college students both by a simple rating scale and by a discrepancy measure. There was a general lack of correlation between these measures and such objective criteria as grade point average, health center visits, army adjustment, the Taylor

*MA* scale, and sociometric status. Coopersmith (1959) compared self-esteem as rated by the self with that estimated by observers, using children as subjects. He suggests that there are actually four types of self-esteem: what a person purports to have, what he really has, what he displays, and what others believe he has.

There is no obvious explanation for the discrepancy of results in studies purporting to relate self-concept to behavioral adjustment. Since the basis for selecting items for rating scales and *Q* sorts differs from study to study, it is possible that the statements used in the scales of those studies with positive results had more of a relationship to the criteria than the statements in studies which were negative.

A different approach to relating self-concept measures to adjustment is shown in a block of psychotherapy research studies at the University of Chicago (Rogers & Dymond, 1954). Change in self-concept was found to occur as a function of improvement during psychotherapy. Butler and Haigh (1954) had clients make *Q* sorts for self and for ideal self both before therapy and after its completion to test the hypothesis that therapy will increase satisfaction with the self. Congruence between the two sorts increased as a result of psychotherapy, the two sorts moving towards a common mean. Rudikoff (1954), using the same subjects, found changes during periods of time before and after therapy were not nearly as great as those occurring during therapy. Also with the same subjects, Dymond (1954) found that there was closer agreement after therapy between the way clients sorted the Butler and Haigh *Q* sort

cards, and the way two non-Rogerian clinical psychologists sorted the cards between what the well-adjusted person should say is like him and what is not like him.

The same investigator (Cartwright, 1958) related change in self-concept over therapy to a successful search for identity. She had clients make sortings with Butler and Haigh *Q* sort cards to describe themselves as they saw themselves in relationship to three people of their choice to test the hypothesis that successful therapy increases the consistency of the self-concept which one brings to different social situations. The hypothesis was confirmed.

Ewing (1954) had counselee college students rate a list of traits for self, ideal self, mother, father, counselor, and a culturally approved figure. There was a regression of the ratings toward a common mean in those clients who were estimated to be the most improved in therapy.

Changes in self-ratings over therapy seem certainly to have occurred. But they seem to take place also without psychotherapy. Taylor (1955) devised a *Q* sort divided between positive and negative statements. After subjects made repeated sortings both for self and for ideal self, he concluded that self-introspection without therapy results in increased positiveness of attitude toward the self; that the self and ideal self will draw closer together; and that repeated self-descriptions are accompanied by increased self-consistency. Engel (1959) studied the stability of self-concept in adolescence, and also found a trend towards more positive *Q* sorting over a 2-year period. And finally Dymond herself (1955) found an increased congruence between *Q* sorts for self and for ideal self among sub-

jects waiting for psychotherapy, although ratings of adjustment based on TAT protocols showed no change over the period.

Dymond attributes increased self-ideal-self congruence without psychotherapy as due to the strengthening of neurotic defenses. It might be charged that similar changes during therapy might have the same basis. Dymond also raises the possibility that the sorts can be influenced by the attitude of the therapist towards the client's self. There is in short no complete assurance that the cognitive self-acceptance as measured by the *Q* sort is related to the deeper level of self-integration that client-centered therapy seeks to achieve.

Indirect evidence of change of the self-concept during counseling is provided by studies showing changes of self-estimates. Several studies show that agreement between self-ratings on interests and the ratings of the self by interest inventories increase as a result of counseling (Berdie, 1954; Froehlich, 1954; Johnson, 1953; Singer & Steffre, 1954). The first two of these studies show a moderate increase in accuracy in predicting one's intelligence, but very little improvement in rating the self on measures of personality. One might reason that some parts of the self-concept are peripheral to the core of the self (e.g., interests) and are therefore unstable, while other parts (e.g., personality estimates) are central to the self and are therefore extremely resistant to change.

#### SELF-CONCEPT—SELF-CONSISTENCY

If the self-concept is to have usefulness as a construct it must be shown that it is consistent in a given self. It must be known whether the self-concept is a gestalt that is more than the

sum of different self-regarding attitudes, or whether instead the self-concept is an impossible attempt to generalize different feelings toward unique situations.

One answer to this question is provided by Akeret (1959). He inter-correlated self-ratings on academic values, interpersonal relations, sexual adjustment, and emotional adjustment, achieving differentially positive interrelationships. Emotional adjustment was the best indicator, correlating + .61 with a total corrected for part-whole inflation. While Akeret interpreted his results as suggesting that an individual does not accept or reject himself totally, the results might also be interpreted as suggesting that some areas of self-regard are more central to the self-concept than other areas.

Consistency in the self-concept was found by Martire and Hornberger (1957), who found very great similarities between measures of the actual self, the ideal self, and a socially desirable self. But inconsistency was found by McKenna, Hofstaetter, and O'Connor (1956), who found that one's self ideal differed less from one's close friends than the close friends differed from each other. These investigators concluded by rather involved reasoning that the ideal self is sufficiently differentiated to seek different need satisfactions in different people.

The search for consistency in the self led also to comparing scores on different measures of self-concept. Omwake (1954) compared three scales—the Bills, Phillips, and Berger—which measure acceptance both of self and of others. The scales were in closer agreement as to the degree of acceptance of self than they were as to acceptance of others. Brownfain (1952) found that low ratings of self were related on his scale to the dis-

crepancy between optimistic and pessimistic self-ratings, or what Brownfain termed stability of self-concept; and Cowen (1954) found a relationship between the pessimistic Brownfain self-ratings, and the discrepancy between self- and ideal-self-ratings on the Bills index. Bendig and Hoffman (1957) found that Bills' scores on acceptance of self-ratings and on congruence between ratings of self and ideal self related equally well to scales of the Maudsley Personality Inventory. They therefore concluded that the two different Bills index measures are redundant.

But on the negative side, Cowen (1956) found no relation between the so-called stability of self-concept on the Brownfain, and the different measures on the Bills. Hampton (1955) likewise failed to find any significant relationship between ability to make realistic appraisals about oneself and the ability to admit statements that were damaging but probably true.

Different measures of the self-concept have different theoretical and operational bases. Where measures apply similar rationale, significant correlations between measures have been found. But in similar measures such extraneous variables as response set and social desirability will produce similar bias. Measures of self-concept have reliability, and in a certain degree are interchangeable. Whether or not the reasons for similarity are intrinsic to the scales, the notion of the internal frame of reference seems well validated.

#### DISCUSSION

The scientist can not hold truths to be self-evident. What is known of the self through direct report must be considered suspect due to philosophical considerations, since the nature

of the "I" has been seen differently in each ideological epoch. Notions concerning the self are like other human ideas, and are inventions and not discoveries. The task is not that of discovering the "true self," but instead of constructing those notions which increase understanding of human behavior. Just as the number of inventions is potentially unlimited, so there need be no limit on the number of constructions put upon the self. In this discussion we will proceed functionally, and consider the uses to which different selves have been put.

The first self is the knowing self of structural psychology. Its function is to apprehend reality. The rational nature of man has always been in dispute, and the New Look in perception has further undermined this conception. This article has cited studies which throw doubt on the ability of the self to perceive itself correctly in those areas which are of great value to it. It is the change in the self as perceiver of itself that is the aim of client centered therapy. Studies of client centered therapy do not reveal whether therapy brings the client any closer to reality, but they do provide some evidence that the perception of the self is brought closer to social expectancies.

The second construction of the self is that of motivator. This is the self of thinkers who believe that the individual is motivated by a need for self-assertion, or self-realization, by realizing those potentialities which inhere within the self. Attempts to validate this construct of the self have been carried on through work on *need achievement*. This construct of the self seems involved also in ratings and *Q* sorts for an ideal self which out-distances the real self. Here, of course, the self whose reach exceeds its grasp is considered to be patho-

logical, for it is shown how psychotherapy helps reduce the disparity between the real and ideal.

The third construct of self is the humanistic, semireligious conception of the self as that which experiences itself. It is the "unique personal experience" of Moustakas (1957) and the experience of feeling in Rogers (1951). The difficulty for the psychologist is that such a conception is more religious than scientific; it becomes a value-orientation, and, as the writer has shown elsewhere (Lowe, 1959), it becomes a highly controversial statement of what is the highest good.

The fourth approach views the self as organizer. This self is the psychoanalytic ego; the internal frame of reference of Snygg and Combs (1949); and the source of construct making in G. A. Kelly (1955). Any operational measure of self-consistency would seem to imply the existence of such a self. It is this self that this article has been most directly concerned with; to the extent that studies have been positive, the self does respond the same way in different situations. Conversely, to the extent that the studies have had negative results there is enough inconsistency in the self that it does not always act according to prediction.

A fifth approach constructs the self as a pacifier. Such a self seems implied in Lewin (1936), who constructed his system of personality in terms of valences or tensions which the organism seeks to keep to a minimum. It seems present also in Angyal (1941) who views life as an oscillation about a position of equilibrium. The self in other words is seen as an adjustment mechanism which seeks to maintain congruence between the self and the nonself. It is the verification of this type of self that seems implied by *Q* sort studies

that show increased congruence of real and ideal self as a result of psychotherapy. We must however note that the self as pacifier stands in direct opposition to the self as motivator.

In the sixth view of the self, the self is the subjective voice of the culture, being purely a social agent. It is the self of both sociology and S-R psychology, for it sees behavioral responses solely in terms of social conditions or stimuli inputs. The self as an entity is denied, and behavioral consistency is seen as residing not in the individual but in similar environmental events. If the term self is used, it is seen in terms of ego-involvements with loyalties which are determinative of the self.

From these different conceptions of the self, we can choose the one which best fits our theoretical frame of reference. But which conception is chosen seems to depend more upon faith than upon logic, and the choice of one conception must of necessity deny other constructs. It seems impossible that the self can function as a motivator which constantly tries to change the status quo, and as a pacifier which minimizes the disparity between the real and ideal self. There is a contradiction also between the self as motivator and the self as feeling, for in the latter the self is accepted as it is, but in the former is not. Differences are apparent also between the self as feeling and as pacifier. And finally, the self as agent of society is opposed to all other conceptions.

## CONCLUSION

Is the self-concept a fact which, having an objective existence in nature, is observed and measured; or is it an epiphenomenon of deeper reality, invented by man that he might better study his behavior?

The world has sought to be so sure of the self because there is so little else of which it can be certain. The self has become the anchor that man hopes will hold in the ebbtide of social change. But just as a fish could never know it was surrounded by water unless that water were to disappear, it is unlikely that Lecky (1945) would have known about self-consistency had he not lived in a culture which felt inconsistency. In Buberian terminology, the self is an It, which man invents because he can not find a Thou.

The position of this paper must be that the self is an artifact which is invented to explain experience. If the self-concept is a tool, it must be well designed and constructed. We will conclude therefore with that construct of the self which best serves the 1960s. Such a construction combines the self of ego-involvement with the self of feeling. It is a self which is existential not to experience itself, but to mediate encounter between the organism and what is beyond. Such a self is what Pfuetze (1954) calls the "self-other dialogic theory of the self," being interpreted naturalistically through Mead and transcendentally through Buber. It is as an artifact that the self-concept finds meaning.

## REFERENCES

AKERET, R. U. Inter-relationships among various dimensions of the self-concept. *J. counsel. Psychol.*, 1959, 6, 199-201.

ALLPORT, G. W. The ego in contemporary psychology. *Psychol. Rev.*, 1943, 50, 451-478.

ANGYAL, A. *Foundations for a science of personality*. New York: Commonwealth, 1941.

BENDIG, A. W., & HOFFMAN, J. L. Bills Index of Adjustment and the Maudsley Personality Inventory. *Psychol. Rep.*, 1957, 3, 507.

BERDIE, R. F. Changes in self-ratings as a

method of evaluating counseling. *J. counsel. Psychol.*, 1954, 1, 49-54.

BERGER, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, 47, 778-782.

BERTOCCI, P. A. The psychological self, the ego, and personality. *Psychol. Rev.*, 1945, 52, 91-99.

BILLS, R. E. Rorschach characteristics of persons scoring high and low in acceptance of self. *J. consult. Psychol.*, 1953, 17, 36-38. (a)

BILLS, R. E. A validation of changes in scores for the Index of Adjustment and Values as measures of changes in emotionality. *J. consult. Psychol.*, 1953, 17, 135-138. (b)

BILLS, R. E. Self-concepts and Rorschach signs of depression. *J. consult. Psychol.*, 1954, 18, 135-137.

BILLS, R. E. *Manual for the Index of Adjustment and Values*. Auburn: Alabama Polytechnic Inst., 1958.

BILLS, R. E., VANCE, E. L., & MCLEAN, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, 15, 257-261.

BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, 51, 254-259.

BROWNFAIN, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, 47, 597-606.

BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts consequent upon client-centered counseling. In C. R. Rogers & R. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 55-75.

CALVIN, A. D., & HOLTZMAN, W. H. Adjustment and discrepancy between self-concept and inferred self. *J. consult. Psychol.*, 1953, 17, 39-44.

CARTWRIGHT, R. D. Effects of psychotherapy on consistency. *J. counsel. Psychol.*, 1958, 4, 15-21.

CHASE, P. H. Self-concept in adjusted and maladjusted hospital patients. *J. consult. Psychol.*, 1957, 21, 495-497.

CHEIN, I. The awareness of self and the structure of the ego. *Psychol. Rev.*, 1944, 51, 304-314.

CHODORKOFF, B. Adjustment and the discrepancy between the perceived and ideal self. *J. clin. Psychol.*, 1954, 10, 266-268. (a)

CHODORKOFF, B. Self-perception, perceptual defense, and adjustment. *J. abnorm. soc. Psychol.*, 1954, 49, 508-512. (b)

COOPERSMITH, S. A method for determining types of self-esteem. *J. abnorm. soc. Psychol.*, 1959, 59, 87-94.

COWEN, E. L. The "negative self-concept" as a personality measure. *J. consult. Psychol.*, 1954, 18, 138-142.

COWEN, E. L. Investigation between two measures of self-regarding attitudes. *J. clin. Psychol.*, 1956, 12, 156-160.

COWEN, E. L., & TONGAS, P. N. The social desirability of trait descriptive terms: Applications to a self-concept inventory. *J. consult. Psychol.*, 1959, 23, 361-365.

DYMOND, R. F. Adjustment changes over therapy from Thematic Apperception Test ratings. In C. R. Rogers & R. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 109-120.

DYMOND, R. F. Adjustment changes in the absence of psychotherapy. *J. consult. Psychol.*, 1955, 19, 103-107.

EASTMAN, D. Self-acceptance and marital adjustment. *J. consult. Psychol.*, 1958, 22, 95-99.

EDELSON, M., & JONES, A. E. Operational exploration of the conceptual self-system and of the interaction between frames of reference. *Genet. psychol. Monogr.*, 1954, 50, 43-140.

EDWARDS, A. L. *Social desirability variables in personality assessment and research*. New York: Dresden, 1957.

ENGEL, M. The stability of the self-concept in adolescence. *J. abnorm. soc. Psychol.*, 1959, 58, 211-215.

EWING, T. N. Changes in attitude during counseling. *J. counsel. Psychol.*, 1954, 1, 232-239.

FIEDLER, R. E., DODGE, JOAN S., JONES, R. E., & HUTCHINS, E. B. Interrelations among measures of personality adjustment in non-clinical populations. *J. abnorm. soc. Psychol.*, 1958, 56, 345-351.

FRIEDMAN, I. Phenomenal, ideal, and projected conceptions of the self. *J. abnorm. soc. Psychol.*, 1955, 51, 611-615.

FROELICH, C. P. Does test taking change self ratings? *Calif. J. educ. Res.*, 1954, 5, 166-169.

HAMPTON, B. J. An investigation of personality characteristics associated with self-adequacy. Unpublished doctoral dissertation, New York University, 1955.

HANLON, T. E., HOFSTAETTER, P. R., & O'CONNOR, J. P. Congruence of self and ideal-self in relation to personality adjustment. *J. consult. Psychol.*, 1954, 18, 215-218.

HILGARD, E. R. Human motives and the con-

cept of the self. *Amer. Psychologist*, 1949, 4, 374-382.

HOLT, R. R. Accuracy of self-evaluations. *J. consult. Psychol.*, 1951, 15, 95-101.

JOHNSON, D. G. Effect of vocational counseling on self-knowledge. *Educ. psychol. Measmt.*, 1953, 13, 330-338.

KELLY, G. A. *Psychology of personal constructs*. New York: Norton, 1955.

KELMAN, H. C., & PARLOFF, M. B. Interrelations among three criteria of improvement in group therapy. *J. abnorm. soc. Psychol.*, 1957, 54, 281-288.

LECKY, P. *Self-consistency*. New York: Island, 1945.

LEWIN, K. *Principles of topological psychology*. New York: McGraw-Hill, 1936.

LOWE, C. M. Value-orientations: An ethical dilemma. *Amer. Psychologist*, 1959, 14, 687-693.

LUNDHOLM, H. Reflections on the nature of the psychological self. *Psychol. Rev.*, 1940, 47, 110-127.

McKENNA, H. V., HOFSTAETTER, P. R., & O'CONNOR, J. P. Concepts of ideal self and of the friend. *J. Pers.*, 1956, 24, 262-279.

MARTIRE, J. G., & HORNBERGER, R. H. Self-congruence by sex and between sexes in a "normal" population. *J. clin. Psychol.*, 1957, 13, 288-291.

MOUSTAKAS, C. *The self*. New York: Harper, 1957.

OMWAKE, K. T. Relation between acceptance of self and acceptance of others shown by three personality inventories. *J. consult. Psychol.*, 1954, 18, 443-446.

PFUETZE, P. E. *The social self*. New York: Bookman, 1954.

PHILLIPS, E. L. Attitudes toward self and others. *J. consult. Psychol.*, 1951, 15, 79-81.

RAIMY, V. C. The self-concept as a factor in counseling and personality organization. Unpublished doctoral dissertation, Ohio State University, 1943.

ROBERTS, G. E. A study of the validity of the Index of Adjustment and Values. *J. consult. Psychol.*, 1952, 16, 302-304.

ROGERS, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.

ROGERS, C. R., & DYMOND, R. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.

ROGERS, C. R., et al. A coordinated research in psychotherapy. *J. consult. Psychol.*, 1949, 13, 149-220.

RUDIKOFF, E. C. A comparative study of the changes in the concepts of the self, the ordinary person, and the ideal in eight cases. In C. R. Rogers & R. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 85-98.

SHEERER, E. J. An analysis of the relationship between acceptance of and respect for self and others. *J. consult. Psychol.*, 1949, 13, 169-175.

SINGER, S. L., & STEFFLRE, B. Analysis of the self-estimate in the evaluation of counseling. *J. counsel. Psychol.*, 1954, 1, 252-255.

SMITH, G. M. Six measures of self-concept discrepancy and instability: Their interrelations, reliability, and relations to other personality measures. *J. consult. Psychol.*, 1958, 22, 101-112.

SNYGG, D., & COMBS, A. *Individual behavior*. New York: Harper, 1949.

STEPHENSON, W. *The study of behavior*. Chicago: Univer. Chicago Press, 1953.

TAYLOR, C., & COMBS, A. Self-acceptance and adjustment. *J. consult. Psychol.*, 1952, 16, 89-91.

TAYLOR, D. M. Changes in self-concept with psychotherapy. *J. consult. Psychol.*, 1955, 19, 205-209.

TURNER, R. H., & VANDERLIPPE, R. H. Self-ideal congruence as an index of adjustment. *J. abnorm. soc. Psychol.*, 1958, 57, 202-206.

ZIMMER, H. Self-acceptance and its relation to conflict. *J. consult. Psychol.*, 1954, 18, 447-449.

ZUCKERMAN, M., & MANASHKIN, I. Self-acceptance and psychopathology. *J. consult. Psychol.*, 1957, 21, 145-148.

(Received May 15, 1960)



GEORGE EASTMAN COMPANY, INC., ROCHESTER, NEW YORK

